WARSAW UNIVERSITY OF TECHNOLOGY

Faculty of Civil Engineering

UNIVERSITY OF WARSAW

College of Inter-faculty Individual Studies
in Mathematics and Natural Sciences

Ph.D. Thesis

Karol Bołbotowski, M.Sc.

Elastic bodies and structures of the optimum form, material distribution and anisotropy

Supervisors:
Professor Tomasz Lewiński, Ph.D., D.Sc.
Professor Piotr Rybka, Ph.D., D.Sc.

Warsaw 2021

Acknowledgements

First I would like to give thanks to my supervisors: Prof. Tomasz Lewiński and Prof. Piotr Rybka. Already in my early student years Prof. Lewiński's lectures made me recognize the beauty of structural mechanics. His exceptional approach to the subject inspired me to pursue a career in research and I have been fortunate to have him as my mentor ever since. To Prof. Rybka I am grateful for his invaluable support on the difficult quest of exploring the more advanced fields of mathematics. Without his lectures and overall guidance it would not have been possible for me to acquire the mathematical tools necessary to complete this thesis. Throughout the long period of my endeavour both supervisors have provided me with countless precious advices: whether on my research, on writing the thesis or on any other responsibility of a PhD student. At the same time they gave me plenty of freedom and space for becoming an independent researcher. Finally, in the more difficult moments, both Prof. Lewiński and Prof. Rybka always offered me a word of comfort that made it easier to carry on. For all of this I cannot be thankful enough.

A great impact on my thesis was made by Prof. Guy Bouchitté, starting with his works on optimal design that pushed me to study mathematics in the first place. I would like to express the deepest of gratitude to Prof. Bouchitté for the opportunity to work with him at Laboratoire IMATH, Université de Toulon for nearly three months at the turn of years 2019 and 2020. Undoubtedly this visit was a turning point for my research and without Prof. Bouchitté this PhD dissertation would not be the same. Our joint work on optimal design of membranes makes up one of the chapter of this thesis.

Many thanks to Prof. Matthew Gilbert for hosting my four-week visit at the Department of Civil and Structural Engineering, University of Shefield in September of 2016. Working under his supervision has changed my perspective on computer methods. In this thesis I take Prof. Gilbert's advice *to sometimes go around the wall instead of climbing it.*

To Prof. Wilfrid Gangbo I am grateful for his kind invitation for a one-month visit at the Department of Mathematics, University of California in Los Angeles in the fall of 2017. Our fruitful discussions have changed my perspective on the way the methods of optimal transport can be engaged in optimal design.

For creating a friendly and stimulating environment I would like to thank the members of the research group led by Prof. Tomasz Lewiński at the Department of Structural Mechanics and Computer Aided Engineering, Faculty of Civil Engineering, Warsaw University of Technology. They are: my bachelor thesis supervisor Prof. Tomasz Sokół, Prof. Grzegorz Dzierżanowski, Prof. Sławomir Czarnecki, Dr Tomasz Łukasiak and Dr Radosław Czubacki. I truly appreciate the inspiring conversations we have had throughout the years.

To the Faculty of Mathematics, Informatics and Mechanics at University of Warsaw I give thanks for the hospitality and the possibility to expand my mathematical horizons. Especially amongst the group of Section of Equations of Mathematical Physics I could feel at home being invited to numerous lectures, seminars, conferences and discussions. Next to one of my supervisors, Prof. Piotr Rybka, I am particularly grateful to Prof. Anna Zatorska-Goldstein.

My special thanks go to Wojciech Górny, an excellent young mathematician and a dear friend of mine. Not once did Wojtek refuse to help me in taking my first steps in measure theory and other fields of mathematics.

Working on this thesis would have been a much greater challenge if it had not been for the huge support of the beloved ones: my Mom, my Dad, my brother Michał and my close friends Daniel, Jakub, Kamil, Lidia, Marcin, Michał, Paulina. Their patience, understanding and persistence in cheering me on were incredible, I will be forever thankful.

Finally, I wish to acknowledge the financial support that I have received throughout the course of my PhD studies. Many thanks to the National Science Centre (Poland) for:

- my research grant no 2015/19/N/ST8/00474 entitled: *Topology optimization of thin elastic shells – a method synthesizing shape and free material design*;
- research grant no 013/11/B/ST8/04436 coordinated by Prof. Tomasz Lewiński and entitled: *Topology optimization of engineering structures. An approach synthesizing the methods of: free material design, composite design and Michell-like trusses*;
- research grant no 2019/33/B/ST8/00325 coordinated by Prof. Tomasz Lewiński and entitled: *Merging the optimum design problems of structural topology and the optimal choice of material characteristics. The theoretical foundations and numerical methods.*

I am also deeply appreciative of the scholarship granted within the Integrated Development Programme (ZIP) at the University of Warsaw.

Abstract

This PhD thesis covers several topics in optimal design of elastic bodies and structures. Its main objective is to employ modern tools of measure theory and duality to establish the mathematical framework of the optimization problems known in literature as well as to propose new ones. The focus is on the specific class of minimum compliance problems that revolve around a pair of mutually dual variational problems: the displacement-based primal problem and its stress-based dual. In the latter an integral functional of linear growth is being minimized rendering the optimal stress field a tensor valued measure – an object capable of modelling multi-dimensional bodies reminiscent of Michell structures.

The first addressed subject is the Free Material Design (FMD) problem originally posed in middle 90's. It consists in finding a field of the Hooke tensor that may freely change from point to point in the design domain. In this thesis the Bouchitté-Buttazzo theory of Mass Optimization Problem was successfully adapted towards FMD. In the process the formulation of FMD has undergone a considerable generalization: it allows a choice of a closed convex cone of admissible Hooke tensors and a wide class of non-linear constitutive laws. The finite element method was built around the corresponding pair of mutually dual problems arriving at an efficient conic quadratic programming formulation.

For the second topic a completely new problem of optimal design of plane membranes is put forth. It starts with a proposal of a geometrically non-linear model of membrane that is inspired by the von Kármán's theory of plates. Despite the non-linearity the corresponding pair of mutually dual problems is recovered. Analysis of the primal displacement-based problem is amongst the greatest challenges of the thesis. Owing to the apparatus of maximal monotone maps the compactness result is established. The numerical scheme is developed through the concept of ground structure known from truss optimization, once more a conic quadratic program is identified. Finally, an unexpected link to the Monge-Kantorovich optimal transport problem is exposed: the optimal design of membranes conceals a problem of finding a metric that maximizes the transportation cost.

The last subject of the thesis is optimal form-finding. One seeks an optimal structure that concentrates on a single surface spanning a plane horizontal region. Only compressive stresses are permitted and the gravitational load migrates vertically so that it tracks the surface. This thesis puts forward a new idea of constructing an optimal 3D form-structure based on the 2D solution of the optimal membrane problem. The sought surface becomes the graph of the membrane's deflection while the 3D stress on the surface comes from unprojecting the plane pre-stress field. The forms thus constructed are not only a far reaching generalization of Prager-Rozvany optimal arch-grids, but are actually proved to be optimal amongst all the 3D structures – concentrating on a surface or not – provided that the load is vertically transmissible.

Key words: optimal design, compliance minimization, Free Material Design, Michell structures, optimal membrane, Monge-Kantorovich problem, Prager problem, form-finding.

Streszczenie

Niniejsza rozprawa doktorska obejmuje tematykę optymalnego projektowania ciał i konstrukcji sprężystych. Głównym założeniem jest wykorzystanie nowoczesnych narzędzi teorii miary oraz dualności w celu opracowania matematycznej teorii zagadnień optymalizacji, zarówno tych znanych w literaturze, jak i nowo zaproponowanych. Uwaga poświęcona jest szczególnej klasie zadań minimalizacji podatności, które mogą być sprowadzone do pary wzajemnie dualnych zadań wariacyjnych: pierwotnego zadania przemieszczeniowego oraz dualnego zadania naprężeniowego. Ze względu na liniowy wzrost funkcjonału całkowego minimalizowanego w zadaniu dualnym optymalne pole naprężeń jest tensorową miarą.

Pierwszym poruszonym zagadnieniem jest Projektowanie Materiału z Wolnego Wyboru (ang. Free Material Design, FMD), sformułowane w połowie lat 90-tych. Polega ono na optymalnym doborze pola tensora Hooke'a. W niniejszej rozprawie teoria zadania FMD została opracowana poprzez dostosowanie metod Optymalnego Projektowania Masy Bouchitté-Buttazzo. Ponadto sformułowanie FMD zostało uogólnione poprzez możliwość wyboru stożka dopuszczalnych tensorów Hooke'a oraz nieliniowego związku konstytutywnego. Wokół pary wzajemnie dualnych zadań opracowana została metoda elementów skończonych, która finalnie prowadzi do zadania programowania stożkowego.

Zadanie optymalnego projektowania płaskiej membrany jest propozycją nowego sformułowania, począwszy od geometrycznie nieliniowego modelu membrany. Pomimo nieliniowości zadanie optymalnej membrany udaje się powiązać z parą wzajemnie dualnych zadań. Analiza przemieszczeniowego zadania stanowi jedno z większych wyzwań rozprawy. Pożądana zwartość zbioru dopuszczalnego została udowodniona dzięki aparatowi maksymalnych funkcji monotonicznych. Metoda numeryczna została zbudowana w oparciu o technikę konstrukcji bazowej znaną z optymalizacji kratownic, po raz kolejny otrzymując zadanie programowania stożkowego. Dodatkowo ukazane zostało niespodziewane połączenie z zadaniem optymalnego transportu Monge'a-Kantorowicza: projektowanie optymalnej membrany jest związane z poszukiwaniem metryki, która maksymalizuje koszt transportu.

Jako ostatnie podjęte zostało zagadnienie optymalnego projektowania przekryć. Poszukiwana jest konstrukcja skupiona na pojedynczej powierzchni rozpiętej nad płaskim poziomym obszarem. Dopuszczalne są wyłącznie naprężenia ściskające, podczas gdy obciążenia grawitacyjne migrują pionowo śledząc powierzchnię. W niniejszej rozprawie zaproponowana jest nowa metoda konstruowania optymalnych przekryć na podstawie rozwiązania dwuwymiarowego zadania optymalnej membrany. Poszukiwana powierzchnia jest wykresem ugięcia membrany, natomiast trójwymiarowe pole naprężeń powstaje przez odrzutowanie napięcia w membranie. Tak budowane przekrycia są nie tylko daleko idącym uogólnieniem koncepcji siatek łukowych Pragera-Rozvany'ego, ale również, jak udowodniono, są one optymalne w klasie dowolnych konstrukcji 3D pod warunkiem, że obciążenia są migrujące.

Słowa kluczowe: optymalne projektowanie, minimalizacja podatności, projektowanie anizotropii, konstrukcje Michella, optymalna membrana, zadanie Monge'a-Kantorovicha, zadanie Pragera, projektowanie przekryć.

Contents

Chapter 1. Introduction 13
1.1. Optimal design – literature overview 13
1.2. The objectives and methodology of the thesis 22
Chapter 2. Mathematical preliminaries 33
2.1. Convex analysis and calculus of variations 33
2.2. Measure theory and convex integral functionals 37
2.3. Multifunctions 44
2.4. Maximal monotone maps on $\mathbb{R}^d$ 47
2.5. The μ-tangential calculus 49
2.6. Conic programming 59
Chapter 3. Including the choice of anisotropy in the optimal design 61
3.1. The energy potential parametrized by the anisotropy 61
3.2. Designing anisotropy at a point 65
3.3. Case study: examples of anisotropy design settings 70
3.4. Constitutive fields as tensor valued measures and the energy integral functional . 86
Chapter 4. The family of Free Material Design problems 93
4.1. Formulation of the Free Material Design problem 93
4.2. The passage to a pair of convex problems $(\mathcal{P})$ and $(\mathcal{P}^*)$. Constructing solution of the (FMD) problem 94
4.3. Relaxation of the problem $(\mathcal{P})$ 99
4.4. Optimality conditions via the μ-tangential calculus 101
4.5. The three main frameworks of the Free Material Design problem 108
4.6. A finite element approximation of the pair $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$ 122
4.7. Michell problem and related topics 139
4.8. Monge-Kantorovich approach to optimal design of heat conductor 149
4.9. Isotropic Masonry Material Design – discussion 158
Chapter 5. Optimal design of plane membranes 163
5.1. The plastic design of the pre-stress field in membranes (the linear model) 163
5.2. Optimal design of elastic membranes (the non-linear model) 167
5.3. Relaxation of the problem $(\mathscr{P})$ 183
5.4. Optimality conditions under extra regularity assumptions 197
5.5. Optimal membranes that are systems of strings. The numerical method by ground structure approach 206

5.6. Maximal Monge-Kantorovich metrics versus optimal design of membranes 242
5.7. Open problems 268

Chapter 6. Optimal design of forms 279

6.1. Formulation of the form-finding problem 279
6.2. Prager problem as a relaxation of the form-finding problem 283
6.3. Optimal forms constructed from optimal membranes 293
6.4. Optimal grid-shells constructed from optimal string systems 305
6.5. Numerical approximation of optimal forms by grid-shells 314
6.6. Final remarks and open problems 321

Bibliography 331

Basic terminology and notation

General topology (X is any topological space):

$\mathscr{U}, \mathscr{V}, \mathscr{W}$	open sets in X
$\overline{A}$ or $\mathrm{cl}(A)$	closure of a set $A \subset X$
$\mathrm{int}(A)$	interior of a set $A \subset X$
∂A	boundary $\overline{A} \setminus \mathrm{int}(A)$ of a set A
$A \Subset B$	$\overline{A} \subset B$ and $\overline{A}$ is compact

Sets and finite dimensional spaces:

$\mathbb{N}, \mathbb{N}_+, \mathbb{Z}, \mathbb{R}$	natural non-negative, natural positive, integer, real numbers
$\mathbb{R}_+$, $\mathbb{R}_-$	the set of non-negative and non-positive real numbers
$\overline{\mathbb{R}}$	the extended real line $[-\infty, \infty]$
$\mathbb{R}^d$, S^{d-1}	d-dimensional Euclidean space and its unit sphere
$\mathbb{R}^{d\times n}$, $\mathcal{S}^{d\times d}$	the space of $d \times n$ matrices and of symmetric $d \times d$ matrices
$\mathcal{S}^{d\times d}_+$, $\mathcal{S}^{d\times d}_-$	the cones of positive and negative semi-definite sym. $d \times d$ matrices
$x \otimes y$	the tensor product of vectors $x \in \mathbb{R}^d, y \in \mathbb{R}^n$, element of $\mathbb{R}^{d\times n}$
$x \overset{s}{\otimes} y$	the symmetric part of $x \otimes y$ when $x, y \in \mathbb{R}^d$, element of $\mathcal{S}^{d\times d}$
I	identity matrix, element of $\mathcal{S}^{d\times d}$
$A \succeq B$	for two matrices $A, B \in \mathcal{S}^{d\times d}$ means that $A - B \in \mathcal{S}^{d\times d}_+$
U, V, W	generic Euclidean spaces
$\langle x_1, x_2 \rangle$	the Euclidean inner product for $x_1, x_2 \in V$
$\lvert x \rvert$	the Euclidean norm $(\langle x, x \rangle)^{\frac{1}{2}}$ of a vector $x \in U$
$\mathrm{Ker}\, A$, $\mathrm{Im}\, A$	the kernel and the image of a matrix $A \in \mathbb{R}^{d\times n}$
$\mathscr{L}(V)$	the space of linear symmetric operators from V to V
$\mathscr{L}_+(V)$	the cone in $\mathscr{L}(V)$ consisting of positive semi-definite operators
Id	identity operator in $\mathscr{L}(V)$
$B(x_0, r)$, $\overline{B}(x_0, r)$	open and closed ball centred at a point $x_0 \in V$ and of radius r
Ω	a domain in $\mathbb{R}^d$ – an open connected set
Δ	the diagonal of the product $\mathbb{R}^d \times \mathbb{R}$, i.e. the set $\{(x, x) : x \in \mathbb{R}^d\}$

Function spaces (X is a topological space, $\mathscr{U} \subset \mathbb{R}^d$ is an open set):

$C(X; V)$	V-valued continuous functions over X
$C_c(X; V)$	functions $f \in C(X; V)$ of compact support in X, i.e. $\mathrm{spt}\, f \Subset X$
$C_0(X; V)$	closure in the ∞ norm of $C_c(X; V)$
$C^{0,\alpha}(\mathscr{U}; V)$	V-valued α-Hölder continuous functions over $\mathscr{U}$, $\alpha \in (0, 1)$ (identified with their unique continuous extensions to $\overline{\mathscr{U}}$)
$C^k(\mathscr{U}; V)$	V-valued functions continuously differentiable over $\mathscr{U}$ up to order $k \in \mathbb{N} \cup \{\infty\}$
$C^k(\overline{\mathscr{U}}; V)$	functions as above whose each of k derivatives extend continuously to $\overline{\mathscr{U}}$

$\mathcal{D}(\mathscr{U};V)$	functions $f \in C^\infty(\mathscr{U};V)$ of compact support in $\mathscr{U}$, i.e. $\operatorname{spt} f \Subset \mathscr{U}$
$\mathcal{D}'(\mathscr{U};V)$	the space of V-valued distributions on $\mathscr{U}$
$\operatorname{Lip}(\mathscr{U};V)$	V-valued Lipschitz continuous functions over $\mathscr{U}$ (identified with their unique continuous extensions to $\overline{\mathscr{U}}$)
$\operatorname{Lip}_0(\mathscr{U};V)$	the intersection $\operatorname{Lip}(\mathscr{U};V) \cap C_0(\mathscr{U};V)$
$L^p_\mu(X;V)$	Lebesgue space of p-integrable V-valued functions with respect to a measure $\mu \in \mathcal{M}_+(X)$ and for $p \in [1,\infty]$
$L^p(\mathscr{U};V)$	the space $L^p_\mu(\mathscr{U};V)$ for the Lebesgue measure $\mu = \mathcal{L}^d$
$W^{k,p}(\mathscr{U};V)$	Sobolev space of V-valued functions on $\mathscr{U}$ of order $k \in \mathbb{N}_+ \cup \{\infty\}$ and for exponent $p \in [1,\infty]$
$BV(\mathscr{U};V)$	V-valued functions of bounded variation on $\mathscr{U}$
$C(X), C_c(X), \dots$	the spaces $C(X;V), C_c(X;V), \dots$ for $V = \mathbb{R}$

Convex analysis and duality:

X^*	the topological dual of a locally convex linear topological space X
$\langle x, x^* \rangle$	the duality pairing for $(x, x^*) \in X \times X^*$, also denoted by $\langle x, x^* \rangle_{(X,X^*)}$
$\operatorname{dom}(f)$	the effective domain of a convex function $f : X \to \overline{\mathbb{R}}$
$f^* : X^* \to \overline{\mathbb{R}}$	the convex conjugate of f, i.e. $f^*(x^*) = \sup_{x \in X} \big\{ \langle x, x^* \rangle - f(x) \big\}$
$\partial f(x) \subset X^*$	the subdifferential of f at x
$g^0 : X^* \to \overline{\mathbb{R}}$	the polar function of a closed gauge $g : X \to [0,\infty]$, i.e. $g^0(x^*) = \sup_{x \in X} \big\{ \langle x, x^* \rangle \,:\, g(x) \le 1 \big\}$
$\chi_A : X \to \overline{\mathbb{R}}$	the indicator function of a set $A \subset X$
$\chi^*_A : X^* \to \overline{\mathbb{R}}$	the support function of a set $A \subset X$
$\Lambda^* : Y^* \to X^*$	the adjoint of a linear continuous operator $\Lambda : X \to Y$
$\operatorname{co}(A)$	the convex hull of a set $A \subset X$

Measure theory (X is a locally compact separable metric space):

Σ, $\mathscr{B}(X)$	a generic σ-algebra and the Borel σ-algebra on X
Σ_μ, $\mathscr{B}_\mu(X)$	the μ-completion of σ-algebras Σ, $\mathscr{B}(X)$
$\mathcal{M}_+(X)$	the set of non-negative Radon measures on X
$\mathcal{M}^b_+(X)$	the set of finite non-negative Radon measures on X
$\mathcal{M}(X;V)$	the Banach space of V-valued Radon measures on compact space X
$\mathcal{M}^b(X;V)$	the Banach space of finite V-valued Radon measures on X (equal to $\mathcal{M}(X;V)$ if X is compact)
$\mathcal{M}^b(X;\mathcal{K})$	the weakly-* closed subset of $\mathcal{M}^b(X;V)$ whose elements are valued in a closed convex cone $\mathcal{K} \subset V$
$\mathcal{L}^d$	the Lebesgue measure on $\mathbb{R}^d$
$\mathcal{H}^k$	the k-dimensional Hausdorff measure on $\mathbb{R}^d$
δ_x	Dirac delta measure at a point x
$\vartheta \ll \mu$	absolute continuity of ϑ with respect to μ
$\vartheta_1 \perp \vartheta_2$	orthogonality of measures

ϑ_a, ϑ_s	the absolutely continuous part and the singular part of measure ϑ with respect to a measure μ (in most cases $\mu = \mathcal{L}^d$)
$\frac{d\vartheta}{d\mu}$	the Radon-Nikodym derivative of $\vartheta \in \mathcal{M}^b(X;V)$ with respect to $\mu \in \mathcal{M}_+(X)$, an element of $L^1_\mu(X;V)$
ϑ_+, ϑ_-	the unique Jordan decomoposition of $\vartheta \in \mathcal{M}^b(X;\mathbb{R})$
$\lvert\vartheta\rvert$	the variation measure of $\vartheta \in \mathcal{M}^b(X;V)$ with respect to the Euclidean norm $\lvert\cdot\rvert$, an element of $\mathcal{M}^b_+(X)$
$\mu \llcorner A$	restriction of measure μ to a μ-measurable set $A \subset X$
$p \# \mu$	the push-forward of measure μ by function p
$\operatorname{spt}\mu$	the support of measure μ

Miscellaneous ($\mathscr{U} \subset \mathbb{R}^d$ is an open set):

$\lVert f\rVert_{\infty,X}$	the infinity norm $\sup_{x\in X}\lvert f(x)\rvert$ for $f \in C(X;V)$
$\operatorname{spt} f$	support of a function $f : X \to V$, i.e. $\operatorname{cl}\big(\{x \in X : f(x) \neq 0\}\big)$
∇u	the gradient of a function $u \in C^1(\mathscr{U};\mathbb{R}^m)$, a function in $C(\mathscr{U};\mathbb{R}^{d\times m})$
$e(u)$	the symmetric part of ∇u for $u \in C^1(\mathscr{U};\mathbb{R}^d)$, a function in $C(\mathscr{U};\mathcal{S}^{d\times d})$
$\nabla^2 u$	the Hessian of $u \in C^2(\mathscr{U};\mathbb{R})$, a function in $C(\mathscr{U};\mathcal{S}^{d\times d})$
$\operatorname{div}\vartheta$, $\operatorname{Div}\varsigma$	distributional divergences of measures $\vartheta \in \mathcal{M}^b(\mathscr{U};\mathbb{R}^d)$ and $\varsigma \in \mathcal{M}^b(\mathscr{U};\mathcal{S}^{d\times d})$, elements of $\mathcal{D}'(\mathscr{U})$ and $\mathcal{D}'(\mathscr{U};\mathbb{R}^d)$ respectively
Du	distributional gradient of $u \in BV(\mathscr{U};\mathbb{R}^m)$, element of $\mathcal{M}^b(\mathscr{U};\mathbb{R}^{d\times m})$
$\varepsilon(u)$	the symmetric part of Du for $u \in BV(\mathscr{U};\mathbb{R}^d)$, element of $\mathcal{M}^b(\mathscr{U};\mathcal{S}^{d\times d})$
$D_H(A,B)$	the Hausdorff distance between subsets A, B of a metric space (X,d)
$d_B(x)$	the distance $\inf_{y\in B} d(x,y)$ between $x \in (X,d)$ and a set $B \subset (X,d)$
$\Gamma : X \to 2^Y$	a multifunction between space X, Y, namely $\Gamma(x) \subset Y$ for every $x \in X$
$\mathcal{G}_\Gamma$	the graph of a multifunction Γ
$\mathbf{v} : \mathbb{R}^d \to 2^{\mathbb{R}^d}$	a maximal monotone map on $\mathbb{R}^d$, a multifunction
$\mathrm{id} : \mathbb{R}^d \to \mathbb{R}^d$	identity map on $\mathbb{R}^d$, i.e. $\mathrm{id}(x) = x$
$\rightharpoonup$	weak convergence
$\overset{*}{\rightharpoonup}$	weak-* convergence

Chapter 1

Introduction

1.1. Optimal design – literature overview

In solid and structural mechanics by an optimal design one understands a body or a structure that achieves the biggest "strength" possible under a constraint on the material's volume. It is classical to assess the body's strength by computing its *compliance*, being a measure of deformation under a given system of loads. The smaller the compliance the higher the strength or the *stiffness*. The history of *compliance minimization problems* is very long and it would be impractical, if not impossible, to make a review of all its aspects. The short survey given below is mainly to focus on the topics and results that are the most significant for this thesis. Unavoidably numerous great achievements in structural optimization will not be mentioned.

1.1.1. From shape optimization to the Free Material Design problem

Various formulations of the optimal design problems may differ in one fundamental matter: the choice of *design variables*. Having a bounded d-dimensional *design domain* Ω one of the most natural approaches is to look for a *shape* – a subdomain $\omega \subset \Omega$ that is to be occupied by a homogeneous isotropic material. The volume fraction $|\omega|/|\Omega|$ is prescribed by a constant $\alpha_0 \in (0, 1)$. One can study sensitivity of the compliance functional with respect to the change of ω, see [161], [44] on the *shape derivative*. A more general approach is not to imply any topology of ω and to improve the design by putting infinitesimal holes or cavities in strategic places. This methodology received a modern mathematical framework known as the *topological derivative*, see [132, 131]. A somewhat opposite technique lies at the basis of the popular numerical *level-set method* (see [6]) where many small holes are initially introduced that can then change shape, merge or vanish.

It has been long recognized (see [111]) that the process of putting increasing number of holes whose size goes to zero may not stop when one attempts to reach the minimum compliance. In other words an optimal "black and white" design is not attainable in general. Plenty of numerical methods in optimal design are precisely aimed at forcing the optimizer to converge to such a desired design. Amongst them the most successful is the SIMP method, see [17, 18]. Although its mathematical correctness is a subject of debate (see e.g. [89]) the impact of the SIMP method on the industrial application of optimal design is unquestionable, cf. [1]. The mathematical take on this issue is different:

the original problem of seeking ω is *relaxed by the homogenization method*, see [4]. The design ω is allowed to partition into two regions: the "black region" ω_1 occupied by the initial material and the "grey region" ω_2 at whose points x the material mixes with the "void" at some ratio $\theta(x) \in (0,1)$ representing density of the homogenized material. The volume constraint then reads: $|\omega_1| + \int_{\omega_2} \theta \, d\mathcal{L}^d = \alpha_0 |\Omega|$. The existence of solution of such a relaxed minimum compliance problem follows. Some of the first numerical methods based on the relaxation by homogenization method may be found in [95], [16].

At points x of the homogenized region ω_2 the composite material is in general anisotropic and it is represented by the *stiffness tensor* $\mathscr{H}(x)$. In this thesis $\mathscr{H}(x)$ will be termed *Hooke tensor* but it will be treated as a linear symmetric positive semi-definite operator from $\mathcal{S}^{d\times d}$ to $\mathcal{S}^{d\times d}$, the latter being the space of symmetric $d \times d$ matrices, namely $\mathscr{H}(x) \in \mathscr{L}_+(\mathcal{S}^{d\times d})$. The range of feasible Hooke tensors at a point x depends on the material density $\theta(x)$ and it is known as the *G-closure*, see [123]. It is the set of all G-limits of sequences of Hooke fields $\mathscr{H}^\varepsilon$ corresponding to microstructures constructed within the *representative volume element* for the volume fraction $\theta(x)$.

The G-closure is by no means easy to determine (see e.g. [128]), which greatly complicates the optimal design by the homogenization method. In [136], [20] the authors propose a bypass: an optimal design problem in which at every point we can freely choose any Hooke tensor $\mathscr{H}(x) \in \mathscr{L}_+(\mathcal{S}^{d\times d})$. Its realizability through microstructures is no longer relevant and we may let go of the homogenization theory completely. Up to a multiplicative constant the trace of the operator $\mathrm{Tr}\big(\mathscr{H}(x)\big)$ is chosen to stand for the material's density $\theta(x)$. This abstract approach is known as the *Free Material Design* (or the *Free Material Optimization*) and its formulation may be written as

$$\inf\left\{\mathcal{C}_1(\mathscr{H}) \;:\; \mathscr{H} \in L^1\big(\Omega; \mathscr{L}_+(\mathcal{S}^{d\times d})\big),\ \int_\Omega \mathrm{Tr}\mathscr{H} \, d\mathcal{L}^d \le C_0\right\} \tag{1.1}$$

where $\mathcal{C}_1(\mathscr{H})$ is the compliance of the design $\mathscr{H}$. Classically it is defined as $\mathscr{F}(\check{u})$, i.e. the virtual work of the load on the deformation $\check{u} : \Omega \to \mathbb{R}^d$ of the elastic body of Hooke tensor field $\mathscr{H}$. If $F \in L^1(\Omega;\mathbb{R}^d)$ are the body forces and $p \in L^1_{\mathcal{H}^{d-1}}(\partial\Omega;\mathbb{R}^d)$ is the boundary loading, the virtual work reads $\mathscr{F}(u) = \int_\Omega \langle u, F\rangle \, d\mathcal{L}^d + \int_{\partial\Omega} \langle u, p\rangle \, d\mathcal{H}^{d-1}$. The deformation function $\check{u}$ is a solution of the PDE system of linear elasticity. However, if $\mathscr{H}$ is degenerate (does not satisfy the standard ellipticity conditions) the notion of such solution complicates as it may lie beyond the Sobolev space $W^{1,2}(\Omega;\mathbb{R}^d)$ or not exist at all. To circumvent these delicate issues it is convenient to define the compliance variationally instead: as the minus total potential energy of the elastic system, namely

$$\mathcal{C}_1(\mathscr{H}) := -\inf\left\{\frac{1}{2}\int_\Omega \big\langle \mathscr{H}e(u), e(u)\big\rangle \, d\mathcal{L}^d - \mathscr{F}(u) \;:\; u \in C^1(\overline{\Omega};\mathbb{R}^d),\ u = 0 \text{ on } \Upsilon_0\right\} \tag{1.2}$$

where $\Upsilon_0 \subset \partial\Omega$ is the Dirichlet part of the boundary and $e(u)$ is the symmetric part of the gradient ∇u. Solution of this minimization problem, should it exist (possibly in some relaxed sense), is the elastic deformation $\check{u}$ and it holds that $\mathcal{C}_1(\mathscr{H}) = \frac{1}{2}\mathscr{F}(\check{u})$.

The Free Material Design (FMD) method can be considered an abstract and far reaching relaxation of the homogenization method: not only the direction of $\mathscr{H}(x)$ can be freely chosen but also the material density $\theta(x) = \mathrm{Tr}\big(\mathscr{H}(x)\big)$ may be arbitrarily large. In

a large portion of works on FMD an upper bound $\bar{\theta} < \infty$ is enforced – mostly for either computational purposes (when a lower bound $\underline{\theta} > 0$ is usually imposed as well) or to gain compactness or both, see e.g. [173], [170], [110], [102], [168, 169]. Further we will be concentrating on the idealized formulation (1.1) with no local bounds on $\mathrm{Tr}\mathscr{H}$.

1.1.2. Optimal design of bar structures and Michell problem

Instead of designing a continuous solid body one can address optimization of a *bar structure* – a finite system of 1D elements that connect at joints to constitute a graph in d-dimensional space. In the literature the main focus is put on design of *trusses*, i.e. structures which transfer the loads by means of axial forces: tensile or compressive, see [153, Chapter 4], [2], [118, Sections 2.1-2.3]. A fair amount of attention was paid to optimization of *grillages* – planar system of beams under bending, see [153, Chapter 5], [118, Section 2.4], [26]. Next to the compliance minimization design – also often referred to as the *elastic design* – the approach of *plastic design* is often preferred. The objective is then to minimize the structure's volume whilst minding the normal stresses to lie within admissible bounds $[-\bar{s}_-, \bar{s}_+]$ where $\bar{s}_-, \bar{s}_+ \in \mathbb{R}_+$ are the permissible stresses for compression and tension.

Optimizing trusses of unknown number of bars leads to a rather distressing conclusion which mirrors the issues emerging in the optimal design of shape of solids: by increasing the number of bars whose cross-sectional areas approach zero we keep on decreasing the structural volume and the process goes on forever. Remarkably this phenomenon has been discerned by A. G. M. Michell already at the dawn of the 20th century, see [127]. Even more extraordinary is the Michell's proposal to handle this peculiarity: the optimal structure being the limit of such a sequence of trusses is neither a bar structure nor a continuous solid body. It is in fact a hybrid of the two – a *generalized structure* in which the bars can be curved and non-prismatic. On top of that in [127] we find a kinematical formulation that underlies the theory of – as they would come to be known – *Michell structures*:

$$\sup\Big\{\mathscr{F}(u) \,:\, u \in C^1(\overline{\Omega};\mathbb{R}^d),\quad u = 0 \text{ on } \Upsilon_0,\quad \rho\big(e(u)\big) \leq 1 \ \text{ in } \overline{\Omega}\Big\} \tag{1.3}$$

where $\rho : \mathcal{S}^{d\times d} \to \mathbb{R}_+$ is the spectral norm that returns the maximal absolute value of eigenvalues of a symmetric matrix. If $\hat{u}$ solves (1.3) then the principal lines of the *generalized stress field* in the optimal structure ought to follow the principal lines of $e(\hat{u})$ provided that the corresponding eigenvalues of the latter are extremal: 1 or -1. First analytical solutions of the *Michell problem* were already established in [127]: the Michell's cantilever supported on a circle and the Michell's half bicycle wheel.

To comprehend the extent to which Michell's work was ahead of its time it suffices to point out that the theory has not progressed significantly until over 50 years later, see [53, 54]. The book by W. S. Hemp [105] is considered a milestone: it puts forth a consistent methodology for finding Michell structures that leads through a system of hyperbolic partial differential equations governing the geometry of the *structural layout* of the optimal frame. Despite the great complexity of this system it has paved the way to finding numerous beautiful analytical solutions of Michell problems throughout the

following decades, see e.g. [53, 54, 55], [114, 115], [98, 99], [159], [164]. The Hemp's method involves manipulating with stress resultants referred to a curvilinear coordinate system – the *Hemp's resultant forces*. An alternative approach to finding the optimal stress field leads through the stress-based problem dual to (1.3). At the formal level it has been derived for the first time in [162]:

$$\inf\left\{\int_\Omega \rho^0(\sigma)\,d\mathcal{L}^d \,:\, \sigma\in L^1(\Omega;\mathcal{S}^{d\times d}),\ \ -\mathrm{Div}\,\sigma = F \ \text{ in } \Omega,\ \sigma\nu_{\partial\Omega} = p \text{ on } \partial\Omega\setminus\Upsilon_0\right\} \quad (1.4)$$

where $\nu_{\partial\Omega}$ stands for the normal to $\partial\Omega$. Above $\rho^0 : \mathcal{S}^{d\times d}\to\mathbb{R}_+$ is the dual of the spectral norm, namely $\rho^0(\sigma) = \sum_{i=1}^d |\lambda_i(\sigma)|$. The minimized integral can be interpreted as the volume of a fully stressed structure.

The groundbreaking ideas of Michell evolved in time to facilitate theories of optimal structures other than trusses. In [141], [120] the authors initiated the research on *optimal grillages*, being in fact plastic plates composed of (possibly) infinitely many beams, see also [142], [133], [153]. Michell's theory then influenced the work [143] on optimal arch-grids. The concept of *optimality conditions* proposed in [134] played an important role in those extensions.

A vast majority of numerical methods developed in the field of optimal design rests upon a finite element formulation, see for instance [17, 4, 168]. The most popular (and at the same time the most natural) numerical scheme for Michell problem – the *ground structure method* – is an exception from this tendency. Put forth for the first time in [75], the method revolves around a truss made of bars interconnecting a "very dense" but finite nodal grid $X\subset\overline{\Omega}$. The cross-sectional areas of the bars then become the design variables and the minimum volume problem gets to be rewritten as a linear programming problem, see [91], [158], [172]. In [32] the ground structure method has been successfully adapted to the optimal grillage problem.

A comprehensive coverage of Michell theory and the aforementioned related topics is to be found in the recent book [118].

At the first glance the two streams of topics in optimal design – shape and anisotropy optimization versus designing generalized bar structures – may seem unrelated. Meanwhile, practically at the same time, in works [15] and [5] a spectacular connection was discovered, see also the later paper [46]. In 2D case the Michell problem can be recast through an asymptotic analysis of optimal shape obtained by homogenization method if we allow the volume fraction α_0 to approach zero. The arguments behind the connection are so far formal – cf. the conjecture through Γ-convergence in [34]. Sadly, the link crumbles in 3D.

The next relation concerns the form of mathematical formulations. In [64] the authors have rewritten the FMD problem (1.1) to a stress-based problem of the form (1.4) known for Michell problem. In the case of FMD, however, ρ^0 is the Euclidean norm $|\cdot|$ on $\mathcal{S}^{d\times d}$ (also known as the Frobenius norm). Similar formulation was already reported in [21], see also the paper [19] where shape sensitivity of stress functionals was studied. If $\hat\sigma$ solves this stress formulation of FMD then, up to a multiplicative constant, the optimal Hooke field $\mathscr{H}$ can be computed as $|\sigma|\,\frac{\sigma}{|\sigma|}\otimes\frac{\sigma}{|\sigma|}$. The dual with respect to the stress-based problem

(or rather the primal in this thesis) is again identical to to the Michell's problem (1.3) up to ρ being the Euclidean norm, see [65] or the earlier conference paper [63].

The reformulation to a pair of the form (1.3), (1.4) has given the FMD problem a new life. Already in [64] one may find the first stress-based finite element method and numerical approximations of optimal Hooke fields $\mathscr{H}$. The same authors have recognized that altering the FMD problem by point-wise restriction of the Hooke tensors to certain subsets $\mathcal{K} \subset \mathscr{L}_+(\mathcal{S}^{d\times d})$ – for instance to the set of isotropic tensors, cf. [62, 68, 67] – does not spoil the passage to the pair (1.3), (1.4) but merely affects the functions ρ, ρ^0. They remain to be two mutually dual norms on $\mathcal{S}^{d\times d}$ but their particular forms are specific to the set $\mathcal{K}$. The word *Free* in the name "Free Material Design" gained a new, deeper meaning. One can speak of different setting of the FMD problem: the *Isotropic Material Design* (IMD) setting, the *Cubic Material Design* (CMD) setting [69], the *Young Modulus Design* (YMD) setting [66].

1.1.3. Measure-theoretic approach to optimal design. The Mass Optimization Problem and the Monge-Kantorovich problem

There is another consequence of the analogies between the Free Material Design and the Michell problem. Solution of the latter lies beyond classical solids or bar structures and so the same can be expected from optimal designs for FMD. Does this find its reflection in the problem (1.4) where the stress σ is a function? This question conveniently brings us to another aspect of optimal design that will be central in this thesis – the mathematical one.

In (1.4) a convex integral functional of linear growth is being minimized. It is coercive in L^1, which results in minimizing over a set that is bounded in this space. Since L^1 is not reflexive it is insufficient for this set to be pre-compact for the weak topology and existence of solution cannot be claimed via the Direct Method of Calculus of Variations. Similar conclusion holds true for the FMD problem (1.1). The well established remedy is to extend the search to the space of Radon measures, in this case of tensor valued measures $\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d})$. Ultimately, Michell problem has waited over a century for its rigorous mathematical formulation through measure theory, which may be found in [42]:

$$\inf\left\{\int_{\overline{\Omega}} \rho^0(\varsigma) \;:\; \varsigma \in \mathcal{M}\big(\overline{\Omega};\mathcal{S}^{d\times d}\big),\quad -\operatorname{Div}\varsigma = F \;\text{ in } \mathcal{D}'(\mathbb{R}^d\setminus\Upsilon_0;\mathbb{R}^d)\right\}. \tag{1.5}$$

Above the divergence is understood in the sense of distributions on an open set $\mathbb{R}^d \setminus \Upsilon_0$, which effectively means that it may be rewritten as the *virtual work principle* for test functions $u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0;\mathbb{R}^d)$. The virtual work of the load is now $\mathscr{F}(u) = \int_{\overline{\Omega}} \langle u, F\rangle$ where $F \in \mathcal{M}(\overline{\Omega};\mathbb{R}^d)$ is a vector valued measure, thus modelling body forces, boundary loads and point loads all at once. Solution $\hat{\varsigma}$ of (1.5) exists as soon as the infimum is finite.

Measures are essential not only from the mathematical point of view: mechanically, they allow to capture the hybrid character of Michell structures that until [42] has been represented through the Hemp's resultant forces, see [118, Chapter 4]. In two dimensions a Michell structure often consists of a "continuous" 2D part and bars that geometrically

are a finite family of C^1 curves $\{\mathrm{C}_i\}_{i=1}^n$. Then the generalized stress field in the Michell structure can be then written as

$$\varsigma = \sigma\,\mathcal{L}^2 \llcorner \Omega + \sum_{i=1}^{n} s_i\,\tau_i \otimes \tau_i\,\mathcal{H}^1 \llcorner \mathrm{C}_i \qquad \text{where } \ \sigma \in L^1(\Omega;\mathcal{S}^{2\times 2}),\ s_i \in L^1_{\mathcal{H}^1}(\mathrm{C}_i;\mathbb{R}) \quad (1.6)$$

where $\tau_i = \tau_i(x)$ is the unit vector tangent to C_i at x. The physical units of ς are the same as of the energy $\int_{\overline{\Omega}} \langle e(u), \varsigma\rangle$, which are $\mathrm{J} = \mathrm{N}\cdot\mathrm{m}$. Accordingly, the stress functions σ and the uniaxial $\sigma_i = s_i\,\tau_i \otimes \tau_i$ are of different units themselves: N/m and N, respectively. In this thesis we shall say that ς is the *absolute stress* while σ, σ_i are *relative stresses* with respect to positive measures $\mathcal{L}^2 \llcorner \Omega$, $\mathcal{H}^1 \llcorner \mathrm{C}_i$ that represent distribution of the material. Those interpretational issues are the price one has to pay for the measure-theoretic approach. Of course there is no guarantee that each solution of Michell problem is of the form (1.6), see [42] for more details.

The idea for representing stress fields as measures is older than the work [42] (which reaches far beyond in other aspects) and can be traced back to the paper [35] on yet another optimal design formulation: the *Mass Optimization Problem* (MOP)

$$\inf\left\{\mathcal{C}_2(\mu)\ :\ \mu \in \mathcal{M}_+(\overline{\Omega}),\ \int d\mu \le C_0\right\}.$$

The compliance functional $\mathcal{C}_2$ is defined similarly as in (1.2) although now the strain energy reads $\int_{\overline{\Omega}} j\big(e(u)\big)d\mu$ where $j : \mathcal{S}^{d\times d} \to \mathbb{R}_+$ is a convex positively p-homogeneous energy potential. It must satisfy an ellipticity condition: $\exists\, C_1 > 0$ such that $C_1|\xi|^p \le j(\xi)$ for all $\xi \in \mathcal{S}^{d\times d}$. For $p = 2$ the potential inducing the linear Hooke law is a possible choice, namely $j(\xi) := \frac{1}{2}\langle H_0\xi, \xi\rangle$ where $H_0 \in \mathscr{L}_+(\mathcal{S}^{d\times d})$ is a constant strictly positive definite Hooke tensor. The measure μ called *mass* in [35] plays the role of an elastic material distribution thus generalizing the notion of material's density θ.

The existence of solution to MOP is showed in [35] under the *balance conditions* for the load F and the supporting zone Υ_0. One of the main achievements of the paper [35], however, is reducing the design problem to the pair of mutually dual programs (1.3) and (1.5). This pair is thus a common factor for problems: the one of Michell, MOP and FMD. In the case of MOP ρ is the unique convex positively one homogeneous function on $\mathcal{S}^{d\times d}$ satisfying $j(\xi) = \frac{1}{p}\big(\rho(\xi)\big)^p$ (ρ is a norm once j is symmetric); the function ρ^0 is the polar of ρ. If the stress $\hat{\varsigma}$ solves (1.5) then the mass $\check{\mu}$ optimal for MOP can be computed as $\rho^0(\hat{\varsigma})$ up to a multiplicative constant. In the paper [35] the authors prove a compactness result for the displacement-based problem (1.3): its relaxed variant admits a solution $\hat{u}$ being an element of $W^{1,q}(\Omega;\mathbb{R}^d)$ for any finite q.

Another challenge that the work [35] on MOP successfully undertakes is writing down a system of *optimality conditions*. In a nutshell these are the conditions for a pair u, ς that are necessary and sufficient for optimality in (1.3), (1.5). If the stress ς is decomposed to $\sigma\mu$ then, aside from the admissibility constraints for the pair of convex problems, the key optimality condition reads $\langle e(u), \sigma\rangle = \rho^0(\sigma)$ μ-a.e. and it is known as the point-wise *extremality condition*. This equation makes sense if the candidate function u is of C^1 class, which is often false in practice. To cover the general case the authors employed the tool

they had developed few years earlier: the *μ-tangential calculus*, see the pioneering work [40]. It involves defining a Sobolev space with respect to an arbitrary finite Radon measure μ denoted by $W^{1,q}_{\mu}$. For its elements u there exists a notion of μ-tangential gradient $\nabla_\mu u$ that is meaningful μ-a.e. Thanks to this beautiful and very technical theory the extremality condition may be rewritten for $e_\mu(u)$ being defined similarly:

$$\langle e_\mu(u), \sigma\rangle = \rho^0(\sigma) \qquad \mu\text{-a.e.}$$

The μ-tangential calculus has been further developed e.g. in [86], [36, 37] and in [150, 27].

Finally, the measure-theoretic approach to optimal design allowed the authors to make yet another connection in the "scalar case", i.e. when $u : \Omega \to \mathbb{R}$ in (1.3) and $e(u)$ is replaced by ∇u. Then, the physical context of MOP is the one of *stationary heat flow* and not of elasticity – we are in fact designing a heat conductor. The dual formulation (1.5) becomes vectorial, namely $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$. In this form it is known in the literature under the name of *Beckmann problem*, see [12]. This form of the pair of mutually dual convex problems opens a path to the famous *Optimal Transport Problem* (OTP) by Monge and Kantorovich, see [166], [152]. The *transportation plan* solving a suitably posed OTP, being a pairing measure $\hat{\gamma} \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$, furnishes another construction of optimal mass $\check{\mu}$ for MOP. Amazingly, in the 2D case there is also a deep link between the Beckmann problem and other renown problem in Calculus of Variations: the *Least Gradient Problem*, see [97, 96, 148, 149].

The very rich theory on the mass optimization problem was revisited in [37] where the approach was proved to be very versatile and flexible. A general framework was put forward in which by specifying a differential operator A one can switch between different physical settings of MOP. For $A = \nabla$ and $A = e$ we readdress the optimal design of heat conductor and elastic body, respectively, while for the Hessian $A = \nabla^2$ we almost automatically enter the new theory of mass optimization in Kirchhoff plates.

1.1.4. The optimal design of forms

We wrap up the literature review by taking a closer look at a design problem that holds a special place in civil engineering and architecture: the problem of *form-finding*. An elastic body that concentrates on a 2D surface S embedded in the three dimensional ambient space is called a *shell*, see [101], [58], [113]. General models of shells assume that they transfer the loads by means of: *membrane forces* and *bending moments*. If the latter action is dominant a relatively bigger thickness of the shell is needed to meet the strength requirements. With the given loading and support conditions it is thus a natural challenge for the designer to seek a shape of the surface S so that bending is eliminated completely. This design process is precisely what we understand by "form-finding". The position of the loads can be either fixed or it can "track" the structure. The latter case often concerns gravitational loads that can migrate vertically.

This approach to structural design can be traced back to Antoni Gaudí (see [22]) who was focused on constructing stone or masonry vaulted forms that carry gravitational loads via pure compression. His works draw attention to another design constraint that may be

implied by the material being used – the one on the sign of the stress. Accordingly, for a canopy made of canvas only tensile membrane stresses are permissible. The great majority of works on finding bending-free shells rely on numerical methods, see e.g. [23], [130]. The method used in a line of papers on the so called *self-supporting structures* stands out, see e.g. [87], [167]. After reducing the problem to the 2D horizontal base plane it employs the concept of *thrust network* developed in [25] – a horizontal network of forces in the in-plane equilibrium is generated based on the Maxwell's idea of reciprocal figures [124] and then suitably erected to constitute a form that carries its own weight by pure compression.

The concept of form-finding does not end at shells: one can also address the problem of forming a 3D truss whose bars do not undergo bending and they lie on a single surface S (Lipschitz continuous at most). In the literature such structures are termed *grid-shells*, which is understandable if the number of bars is high – the steel-glass roof of the shopping mall "Złote Tarasy" in Warsaw can serve as an example. Engineers have long been searching for efficient methods of finding equilibrated configurations of grid-shells, see for instance the *dynamic relaxation method with kinetic damping* in [73]. Up to this day the *Force Density Method* introduced in [154] is one of the most popular: the coordinates of the truss' joints can be determined by solving a linear system of algebraic equations. The latter method allows enforcing the sign of axial forces.

Designing a surface structure that avoids bending is already economically reasonable but it may be treated as a starting point for further improvement – optimization, e.g. in terms of compliance. The optimization can be restricted to finding a material distribution in a membrane shell whose middle surface S is prefixed: see [160] on the shape derivative or [77] on using the homogenization method. A much taller order is to optimize the form S and its material properties all at once. With its inherent non-convexity this problem is mainly a subject of papers on sophisticated numerical methods, cf. [24], [106]. In case of grid-shells the design of optimal forms is usually achieved by combining classical form-finding methods like the Force Density Method with numerical optimization algorithms, see [135], [79].

Amongst various formulations of *optimal form-finding* one has drawn perhaps the most attention. Over a planar horizontal domain $\Omega \subset \mathbb{R}^2$ we are to design an optimal form that by means of pure tension (or pure compression) transfers vertical gravitational loads to the supporting curve $\partial\Omega$. The surface on which the form concentrates can be represented as the graph S_z of an *elevation function* $z \in C(\overline{\Omega}; \mathbb{R})$. The load vertically *migrates* together with the surface S_z. Its intensity is given with respect to the horizontal *reference plane* therefore the loading data is expressed through a scalar measure $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ (acting upwards if positive). If f is finitely supported, e.g. on a set $\{x_i\}_{i=1}^n$, then the ultimate load is a system of vertical point forces whose application points are elevated by $z(x_i)$. The form can be optimized in terms of compliance (more common for shell structures) or of volume when a plastic grid-shell is assumed, see [49, 119].

In the remainder of this section we will be focusing on this specific class of optimal form-finding problems. Its plane variant – where $\Omega = (a, b) \subset \mathbb{R}$ and a curve S_z constituting a least-volume hanging cable in tension (or a masonry arch in compression) is being sought – can be solved analytically as it is directly linked to *the funicular problem*, cf. [70]. With

bending dismissed the cable's equilibrium enforces that the elevation function $z : [a, b] \to \mathbb{R}$ satisfies equation $-s\, z'' = f$ where $s \in \mathbb{R}_+$ is the horizontal pulling force. Owing to the boundary conditions $z(a) = z(b) = 0$ the elevation z is determined up to multiplicative constant $1/s$. Then, assuming that the cable is fully stressed, one can find the optimal force s that minimizes the volume.

In contrast to the funicular, the shape of surface form S which carries the load f via compression is highly non-unique, thus rendering the 3D optimal form-finding problem a much more involved design problem. Nonetheless, the way of constructing optimal funiculars sparks an idea for a general methodology: to reduce the three dimensional optimal form design to a convex optimization problem in the 2D reference domain Ω. The first paper that follows this path is [143] where G. I. N. Rozvany and W. Prager has put forth the theory of *optimal arch-grids* (if compression is assumed and of *optimal cable networks* for the tension-only setting). For a convex domain $\Omega \subset \mathbb{R}^2$ and a Cartesian frame $\mathrm{e}_1, \mathrm{e}_2$ chosen for the reference plane $\mathbb{R}^2$ an arch-grid is a 3D structure that decomposes to two family of arches being curves going from one point of $\partial\Omega$ to another while their projections onto $\mathbb{R}^2$ are segments parallel to e_1 or e_2. The goal is to minimize the structural volume while complying with the stress regime – the work [143] focuses on the plastic design. The exceptionality of this form-finding problem lies in its reformulation to a maximization problem on Ω which is reminiscent of the Michell's problem (1.3):

$$\sup \left\{ \int u \, df \ : \ u \in C^1(\overline{\Omega}; \mathbb{R}),\ u = 0 \text{ on } \partial\Omega, \quad \begin{array}{l} \frac{\partial u}{\partial x_1}(\,\cdot\,, x_2) \in \mathrm{K}_1(x_2) \quad \forall\, x_2, \\ \frac{\partial u}{\partial x_2}(\,\cdot\,, x_1) \in \mathrm{K}_2(x_1) \quad \forall\, x_1 \end{array} \right\}. \tag{1.7}$$

The constraints on the partial derivatives are called the *mean squared slope conditions* and they are complicated to write down explicitly. If for x_2 belonging to the projection of Ω onto the second coordinate we put $\big((x_1^-, x_2), (x_1^+, x_2)\big) = \Omega \cap (\mathbb{R} \times \{x_2\})$ then $\mathrm{K}_1(x_2)$ is the non-empty convex set of univariate functions whose L^2 norm on (x_1^-, x_1^+) is non-greater than $\sqrt{2}\,|x_1^+ - x_1^-|^{\frac{1}{2}}$. It is remarkable that if u solves (1.7) then the optimal configuration of arches concentrates on a single surface being the graph S_z of the function

$$z := -\frac{1}{\sqrt{2}}\, u,$$

whilst the sign above changes if a cable network in tension is to be laid on S_z. The optimal arch-grid problem has been recently revisited in [70] where the dual to (1.7) was derived: it involves minimizing a convex functional over the fields of transverse forces $q : \Omega \to \mathbb{R}^2$ satisfying an out-of-plane equilibrium equation $-\mathrm{div}\, q = f$ in Ω. This duality framework paved a way to numerical methods in [70], see also [78].

The research on optimal arch-grids may be considered one of the very few that build a bridge between form-finding and theory of optimal design. Nevertheless, the constraint on the stress trajectories to follow an orthogonal (in the top-view) network of arches is heavy. A generalizations have been looked for in the literature: not long after W. Prager's passing the authors of the work [146] wrote:

"*However, Prager's original intention was to optimize the layout of the arch centrelines as well as the shape of the middle surface for given boundary conditions and loading.*"

In the very same paper the authors introduced the concept of *Prager structures*. These are generalized 3D structures – possibly composed of solids, surface parts, straight or curved bars – that over a planar region Ω transfer a vertically migrating load f to the supporting curve $\partial\Omega$ with the minimum volume. Again one of the two settings may be considered: the pure-compression or the pure-tension setting. The requirement that the optimal design is a form is now technically gone but the possibility of moving the load, being the next design variable, is to encourage form-like solutions. This is in fact what happens in the 2D case, i.e. a funicular arch (or cable) of optimal rise solves the *Prager problem* in the vertical plane, see the formal proof in [145]. In three dimensions the theory of Prager structures have not gone far. In [146] a proposal of a bar structure that could be a solution of the problem for the load being one force can be found – due to a strong assumption on the configuration of the bars the result must be treated as a conjecture. In the same work solutions for cylindrical symmetric data were constructed by revolving the 2D funicular solutions.

In the more contemporary works the Prager problem is to be found under the term *optimal design for transmissible loads*. A load f is called transmissible if not only it can be moved but also arbitrarily distributed vertically – a point force may be converted into a load smeared along the vertical line. There are numerous works which through numerical simulations point to the fact that when loads are transmissible the 3D solutions tend to concentrate on a single surface. This observation concerns both "continuous" structures, see [88], [56] where FEM is used, and trusses, see [92], [107], [117] that rely on the ground structure approach. In the papers [72, 164] it was showed, both numerically and analytically, that restricting to either pure-compression or pure-tension setting is essential for form-structures to be optimal, see also the study in [122].

To the best knowledge of the present author the question on whether form-structures solve the Prager problem (or a similar optimal design formulation for transmissible loads) has been so far unresolved.

1.2. The objectives and methodology of the thesis

The foregoing literature review distinguishes two groups of optimal design formulations. The first starts with optimizing the shape of a solid and aims at designs that are manufacturable, e.g. by 3D-printing. Its mathematical treatment requires the theory of homogenization and the underlying tools of Γ-convergence. The second approach concentrates on structures in a generalized sense: bodies whose dimensionality and anisotropy may change from point to point, whose material density may vary from zero to infinity. The mathematical burden then lies in the use of measure theory. Although such generalized structures may not always be directly applicable in practice, they can provide crucial information in terms of conceptual design. An engineer or an architect is able to recognize the main layout of the structure that proves to be the most efficient, e.g. in the bearing systems of high-rise buildings, see [163], [125], [121], [31], [104]. On top of that, those

structures help to realize where is the *ideal* that may be then pursued by other methods of optimal design.

The focus of this dissertation is entirely on the second topic – the optimal design of generalized structures. From the point of view of mechanics the thesis will cover several different formulations. Each single one, however, will be effectively brought to a pair mutually dual problems on $\Omega \subset \mathbb{R}^d$:

$$Z := \sup\Big\{ \mathscr{F}(u) \ : \ u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U), \quad \rho(Au) \leq 1 \ \text{ in } \overline{\Omega} \Big\} \tag{$\mathcal{P}$}$$

$$Z = \inf\Big\{ \int_{\overline{\Omega}} \rho^0(\varsigma) \ : \ \varsigma \in \mathcal{M}\big(\overline{\Omega}; V\big), \quad A^*\varsigma = \mathscr{F} \ \text{ in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U) \Big\} \tag{$\mathcal{P}^*$}$$

The actual optimal design problem will be encoded by the set of parameters: finite dimensional spaces U, V; a linear differential operator $A : \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U) \to C_0(\mathbb{R}^d \setminus \Upsilon_0; V)$, a distribution $\mathscr{F} \in \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U)$ and finally a pair of mutually polar functions $\rho, \rho^0 : V \to \overline{\mathbb{R}}$ that are *closed gauges* – non-negative convex positively one-homogeneous and lower semi-continuous extended-real functions.

The contribution of the thesis can be divided into two parts: theoretical and numerical, whereas a greater emphasis is put on the first one. The primary goal of the author was to make the theory fully rigorous in mathematical terms so that the thesis lies at the intersection of two disciplines: mechanics and mathematics. The purpose of the proposed numerical methods is mainly to demonstrate the computational and thus also practical potential of the derived formulations $(\mathcal{P}), (\mathcal{P}^*)$. Their justification in the form of convergence analysis is not a part of this dissertation.

1.2.1. Setting the Free Material Design problem through the methods of Mass Optimization Problem

The first optimal design problem addressed in the thesis is the Free Material Design (FMD) problem. Above all, the main objective is to develop its rigorous mathematical framework. The results are partitioned into two Chapters 3, 4 and in large part they make up the joint work with Tomasz Lewiński, see the preprint [30].

The difficulties start at the level of formulation of the problem so that it is well posed. All the works that has established existence so far (see e.g. [20], [170]) imposed an additional point-wise bound on the trace of the Hooke function $\mathrm{Tr}\big(\mathscr{H}(x)\big) \leq c_0$ so that the minimization runs through a closed ball in $L^\infty\big(\Omega; \mathscr{L}_+(\mathcal{S}^{d\times d})\big)$ where there is sufficient compactness. This way the use of measure theory is bypassed. Herein we wish to dispose of the extra local constraint.

The next goal is to provide the mathematical treatment of the ideas put forth in [64, 65] that translate the Free Material Design to a pair of problems of the type $(\mathcal{P}), (\mathcal{P}^*)$. This includes construction of an optimal field of Hooke tensor based on a solution $\hat{\varsigma}$ of $(\mathcal{P}^*)$.

The methodology for tackling the mathematical challenges outlined above is to be found in the works on Mass Optimization Problem [35, 41, 37] that employ measure theory and duality. In the case of the Free Material Design problem, however, instead of working

with a non-negative measure $\mu \in \mathcal{M}_+(\overline{\Omega})$ representing distribution of the material only, the Hooke field that we seek must be modelled by a tensor valued Radon measure $\lambda \in \mathcal{M}\big(\overline{\Omega}; \mathscr{L}_+(\mathcal{S}^{d\times d})\big)$. Nonetheless each such measure admits a (non-unique) decomposition $\lambda = \mathscr{H}\mu$ where $\mu \in \mathcal{M}_+(\overline{\Omega})$ can be again interpreted as material's distribution while the function $\mathscr{H} \in L^1_\mu\big(\overline{\Omega}; \mathscr{L}_+(\mathcal{S}^{d\times d})\big)$ point-wisely informs of the anisotropy.

The adaptation of the Mass Optimization Problem (MOP) methods towards Free Material Design in its original setting – with the linear constitutive law, the full set of admissible Hooke tensors $\mathscr{L}_+(\mathcal{S}^{d\times d})$ and the material's cost point-wise given as the trace of the tensor – is one challenge, but the scope of this thesis goes beyond that. In [62, 69, 66] one finds other formulations of FMD with restrictions on the set $\mathcal{K}$ of admissible Hooke tensors. The plan is not only to synthesize the FMD theory for a generic set $\mathcal{K}$ but also to treat more general constitutive laws that for each Hooke tensor $H \in \mathcal{K}$ are induced by a potential $j(H, \cdot) : \mathcal{S}^{d\times d} \to \mathbb{R}_+$. On top of that, cost functions $c : \mathcal{K} \to \mathbb{R}_+$ different than Tr are to be allowed. The work done on the Free Material Design will thus exceed the purely mathematical analysis – the thesis also deals with generalization of the problem, both in terms of the material model j and the design setting $\mathcal{K}, c$.

Ultimately a family of Free Material Design problems parameterized by the triple $\mathcal{K}, j, c$ is put forward and studied in this thesis:

$$\mathcal{C}_{\min} = \mathcal{C}_{\min}(\mathcal{K}, j, c) := \inf\left\{\mathcal{C}(\lambda) \ : \ \lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K}),\ \int_{\overline{\Omega}} c(\lambda) \le C_0\right\} \tag{FMD}$$

where the compliance is defined variationally

$$\mathcal{C}(\lambda) = \mathcal{C}(\lambda; j) := \sup\left\{\mathscr{F}(u) - \int_{\overline{\Omega}} j\big(\lambda, Au\big) \ : \ u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U)\right\}.$$

Another parameter that comes into play is the linear differential operator $A : \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U) \to C_0(\mathbb{R}^d \setminus \Upsilon_0; V)$. It is perhaps the most substantial one as it allows to choose different mechanical or, even more generally, physical framework of (FMD) ($A = e$ induces elasticity). This universal approach is not a novelty as it was already proposed for MOP, see [37].

Naturally, suitable conditions must be imposed on the triple $\mathcal{K}, j, c$ so that (FMD) is well posed. We choose to work with arbitrary closed convex cone $\mathcal{K}$ and any cost c that is a restriction of a norm to $\mathcal{K}$. Assumptions on the energy function

$$\mathcal{K} \times V \ni (H, \xi) \ \mapsto \ j(H, \xi) \in \mathbb{R}_+$$

need more attention. Requiring convexity with respect to ξ is standard. One-homogeneity with respect to Hooke tensors H is essential for the integral $\int_{\overline{\Omega}} j\big(\lambda, Au\big)$ to be meaningful. Next, concavity and upper semi-continuity with respect to H must imposed if the compliance mapping $\lambda \mapsto \mathcal{C}(\lambda)$ is to be convex and weakly-* lower semi-continuous. This set of conditions turns out to be enough for the well-posedness of (FMD) but if we wish to make a connection to the pair $(\mathcal{P}), (\mathcal{P}^*)$ then one more is needed: for every $H \in \mathcal{K}$ the potential $j(H, \cdot)$ must be positively p-homogeneous for some $p \in (1, \infty)$.

The passage from (FMD) to the pair $(\mathcal{P}), (\mathcal{P}^*)$ is similar as the one from MOP to this pair, see [35, 37]. The fundamental difference lies in the relation between the gauges ρ, ρ^0

and the energy function j, which in case of MOP was algebraic. In the case of *Anisotropic Material Design* setting, namely for $\mathcal{K} = \mathscr{L}_+(\mathcal{S}^{d\times d})$, $j(H,\xi) = \frac{1}{2}\langle H\xi,\xi\rangle$, $c(H) = \operatorname{Tr} H$, in the work [64] it was shown that both of the functions are Euclidean norms: $\rho = \rho^0 = |\cdot|$. The functions change and are very specific when $\mathcal{K}$ is the set of isotropic tensors, see [62]. The aim of this thesis is to work with the general case when those relations turn out to revolve around finite dimensional programming problems involving the whole triple $\mathcal{K}, j, c$:

$$\frac{1}{p}\big(\rho(\xi)\big)^p = \max_{H\in\mathcal{K},\ c(H)\le 1} j(H,\xi), \qquad \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = \inf_{H\in\mathcal{K},\ c(H)\le 1} j^*(H,\sigma), \tag{1.8}$$

where $p' = \frac{p}{p-1}$ and (by an abuse of notation) we agree that $j^* : \mathcal{K}\times V \to \overline{\mathbb{R}}$ is the convex conjugate of j with respect to the second variable. The two programs require studying, including the structure of $\bar{\mathcal{K}}_1(\xi)$ and $\underline{\mathcal{K}}_1(\sigma)$ being the sets of solutions of the maximization and, respectively, minimization problem above. Especially finding solutions of the latter can be difficult if approached directly, hence we will be looking for a characterization of the set $\underline{\mathcal{K}}_1(\sigma)$ (see Theorem 3.14).

The wide range of possible choices of $\mathcal{K}, j, c$ makes it possible not only to recast the settings of (FMD) already known from the literature [64, 62] but also to propose new ones. This opportunity is seized to put forward the *Fibrous Material Design* setting (FibMD) where $\mathcal{K}$ is the convex hull of the set of uni-axial Hooke tensors:

$$\mathcal{K} = \mathcal{K}_{fib} := \operatorname{co}\bigg(\Big\{a\,\tau\otimes\tau\otimes\tau\otimes\tau\ :\ a\ge 0,\ \tau\in S^{d-1}\Big\}\bigg),$$

see Example 3.18. It turns out to be exceptional since it renders ρ the spectral norm, which in turn makes the pair $(\mathcal{P}), (\mathcal{P}^*)$ coincide with Michell problem – we arrive at another optimal design formulation that next to [5, 15, 46, 34] is equivalent to Michell problem (see Section 4.7), although the connection is less deep. The fairly weak assumptions on the energy function become precious when designing structures made of material that is sensitive to the sign of the stress. Departing from $j(H,\xi) = \frac{1}{2}\langle H\xi,\xi\rangle$ we can construct an energy function $j_+ : H\times\mathcal{S}^{d\times d}\to\mathbb{R}_+$ that satisfies the assumptions and

$$j_+(H,\xi) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+} j(H,\xi+\zeta), \qquad j_+^*(H,\sigma) = j^*(H,\sigma) + \chi_{\mathcal{S}^{d\times d}_+}(\sigma),$$

see Section 3.3.2. The energy j_+ models a *canvas-like material* that cannot withstand compression. Analogously, j_- may be defined to model masonry material. Finally $j_\pm = \alpha_+ j_+ + \alpha_- j_-$ where $\alpha_+, \alpha_- \in (0,\infty)$ furnishes a material that behaves dissymmetrically in tension and compression.

The effectiveness of the method that reduces the Free Material Design problem to the pair $(\mathcal{P}), (\mathcal{P}^*)$ lies in constructing solution λ of (FMD) based on a stress field ς solving $(\mathcal{P}^*)$ (see Theorem 4.6). Once we make a decomposition $\lambda = \mathscr{H}\mu$, computing the elastic material distribution μ is the easy part already known from theory MOP: up to a multiplicative constant it must be equal to $\rho^0(\varsigma)$. Choosing the optimal anisotropy function $\mathscr{H}$ is more subtle. If σ is the relative stress satisfying $\varsigma = \sigma\mu$ there must hold $\mathscr{H}(x)\in\underline{\mathcal{K}}_1\big(\sigma(x)\big)$ for μ-a.e. x. The delicate question is the following: under the chosen assumptions on $\mathcal{K}, j, c$ does there exist such a function $\mathscr{H}$ which is μ-measurable? (see Lemma 3.41)

All in all, the theoretical treatment of the Free Material Design problem put forward in this thesis may be divided into two phases. The first phase, constituting Chapter 3, is somewhat preparatory – it gathers new results on involving the choice of anisotropy in the optimal design. Apart from analysis of the finite dimensional programs (1.8) at an abstract point x (see Section 3.2) and its case study for various $\mathcal{K}, j, c$ (see Section 3.3), an analysis of the integral functional $(\lambda, \epsilon) \mapsto \int j\big(\lambda, \epsilon\big)$ is carried out (see Section 3.4). The second phase is the core of Chapter 4 and may be summarized as adaptation of the methods of Mass Optimization Problem through applying the results from Chapter 3. Unavoidably plenty of the steps in Chapter 4 will be repeated after [35, 37] or slightly altered. Most of the proofs will be included to assure that the thesis is self-contained. In particular, we shall follow [37] in deriving the optimality conditions in the framework of μ-tangential calculus – the new challenge will consist in involving the Hooke function $\mathscr{H}$ (see Theorem 4.22). Similarly in Section 4.8, where we diverge from elasticity to give a link between Free Material Design of heat conductors and the Monge-Kantorovich problem, we generalize the spectacular results of [35].

Most numerical methods for the Free Material Design problem (in its original setting [20]) attack the formulation (FMD) directly, see e.g. [168, 169] for a finite element method using semi-definite programming. The goal here is to promote the computational significance of the pair $(\mathcal{P}), (\mathcal{P}^*)$ and the finite element method will be built around it exactly (see Section 4.6). The FE discretization is performed on the displacement-based problem $(\mathcal{P})$. Using element-wise affine shape functions makes it possible to find a *conic programming* formulation of the finite dimensional problem $\mathbf{P}_h$ thus unlocking efficient interior-point methods [9]. Solution of the dual conic program $\mathbf{P}_h^*$ furnishes an element-wise constant approximation $\sigma_h \in L^1(\Omega; \mathcal{S}^{d\times d})$ of solution to $(\mathscr{P}^*)$ that in turn paves a way to approximating (FMD) with $\mathscr{H} \in L^1(\Omega; \mathcal{K})$.

1.2.2. Optimal forms versus optimal planar membranes

The contribution of this thesis towards the theory of Free Material Design is mostly mathematical – it makes a fusion of the mathematical methods of MOP [35, 37] with recent advances of (FMD) [64, 65, 62], some generalizations are introduced along the way.

A different story may be told about another topic covered in Chapter 6 – the design of optimal forms, see the preprint [28]. The goal is to establish the theory of optimal forms that was outlined by W. Prager and then pursued by G. I. N. Rozvany and his co-workers [146, 145]. In the spirit of their research by an *optimal form* we will understand a fully stressed structure that has a minimum volume in the following class:

- the structure concentrates on a surface S_z spanning a plane region $\Omega \subset \mathbb{R}^2$,
- it carries the load $\mathrm{e}_3 f$ that vertically migrates by tracking S_z,
- it undergoes tensile stresses only (or compressive only), bending moments are excluded,
- *the trajectories of stresses on S_z are free of any constraints*,

where $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ is the load data, e_3 is the vertical unit vector and S_z is a graph of some continuous function $z \in C(\overline{\Omega}; \mathbb{R})$ that vanishes on $\partial\Omega$. The last condition is what makes a

difference between this optimal form-finding formulation and Rozvany-Prager theory of optimal arch-grids. The generalization is by no means straightforward since decomposition of the form to arches is the very foundation of the arch-grid problem. Loosely speaking, the optimal form-finding problem thus posed may be thought of as a two step design process: after choosing the best surface S_z we are designing a Michell-like (the stresses must be of one sign) structure on this surface. This perspective may be useful when developing numerical methods, but does not do much good for the analysis of the problem. It is in fact non-convex with respect to the pair of variables: z and the stress field on S_z, regardless of how one defines the latter. Ultimately, direct treatment of this 3D design problem (see Section 6.1 for precise formulation) appears to be very difficult and the goal is to repeat the feat achieved in the case of arch-grids – to find a well posed 2D formulation on Ω whose solution may be utilized to construct the optimal form.

First let us take a step back, from optimal 3D arch-grids to their source of inspiration: optimal cable in tension (or arch in compression) being a 2D funicular. Constructing the elevation function $z : [a, b] \to \mathbb{R}$ that is optimal for the vertical load $f \in \mathcal{M}([a, b]; \mathbb{R})$ can be reinterpreted. Assume a straight *string* connecting a to b that is pre-stressed by a non-zero force $s \in \mathbb{R}_+$. Then its deflection $u : [a, b] \to \mathbb{R}$ is precisely the solution of the equation $-s\,u'' = f$ appended with Dirchlet conditions $u(a) = u(b) = 0$. The elevation of the optimal funicular can be then found as $z = \alpha u$ (a cable in tension) or $z = -\alpha u$ (an arch in compression) for a suitably chosen $\alpha \in \mathbb{R}_+$ and the force in the optimal funicular can be computed by unprojecting $\frac{1}{\alpha}\, s$ or, respectively, $-\frac{1}{\alpha}\, s$.

A similar reinterpretation can be proposed for constructing optimal arch-grids: instead of a 3D grid of arches we may speak of a 2D network of strings that fills the planar domain Ω. The strings are parallel either to e_1 or e_2, they connect the opposing points of $\partial\Omega$ and are pre-stressed by non-negative forces $s_1 = s_1(x_2)$ and, respectively, $s_2 = s_2(x_1)$. The forces are constant along the strings but may change from one to another. The two families of strings then undergo deflection $u : \overline{\Omega} \to \mathbb{R}$ due to the load $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$. The job (that was trivial for a single string) is to choose the distribution of the forces s_1, s_2 so that the virtual work $\int_{\overline{\Omega}} u\, df$ is minimal under the constraint on "the integral of $s_1 + s_2$".

The next idea is the key one as it has led the author of this thesis to generalization of the theory of optimal arch-grids as was intended by Prager and Rozvany. The foregoing orthogonal families of strings may be considered a particular anisotropic case of a *pre-stressed membrane* modelled by a positive semi-definite planar stress field $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{2\times 2}_+)$. Since the membrane is subject to an out-of-plane loading f only, the in-plane equilibrium implies that $\operatorname{Div}\varsigma = 0$ in the interior Ω of the domain. The orthogonal string network is recast if $\varsigma = s_1\, \mathrm{e}_1 \otimes \mathrm{e}_1 + s_2\, \mathrm{e}_2 \otimes \mathrm{e}_2$ where $s_1, s_2 \in \mathcal{M}_+(\overline{\Omega})$; note that the condition on divergence rules out varying of s_1, s_2 with respect to x_1, x_2, respectively.

The strategy for optimal form-finding can be readily guessed – we must let go of the constraint on the principal directions of ς. For any membrane $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{2\times 2}_+)$ satisfying $\operatorname{Div}\varsigma = 0$ in Ω its compliance can be defined variationally:

$$\mathcal{C}(\varsigma) = \sup\left\{ \int_{\overline{\Omega}} u\, df - \frac{1}{2}\int_{\overline{\Omega}} \langle \nabla u \otimes \nabla u, \varsigma\rangle \;:\; u \in \mathcal{D}(\Omega)\right\}. \tag{1.9}$$

Formally, solution u of the maximization problem is the solution of Euler-Lagrange equation $-\mathrm{div}(\varsigma\nabla u) = f$. For the isotropic and homogeneous field $\varsigma = s\,\mathcal{L}^2 \llcorner \Omega$ we recover the classical membrane model [61, Section IV.10.3] governed by the Poisson equation $-s\,\triangle u = f$. The proposal of the 2D convex optimization problem that shall underlie the design of optimal forms is as follows:

$$\inf\left\{\mathcal{C}(\varsigma)\ :\ \varsigma\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{2\times 2}_{+}),\ \ \mathrm{Div}\,\varsigma = 0\ \ \text{in }\Omega,\ \int_{\overline{\Omega}}\mathrm{Tr}\,\varsigma\le C_0\right\} \tag{OPM}$$

It will be referred to as the *Optimal Plastic Membrane* problem – the word "plastic" ought to reflect the control over the stress ς without minding the constitutive law of elasticity. The recipe for constructing the optimal form in 3D mimics the one for optimal funicular in 2D, but the new substantial part is precisely the optimization of the pre-stress:

$$\begin{cases}\varsigma \text{ solves (OPM)},\\ u \text{ solves (1.9) for } \varsigma\end{cases}\quad\Rightarrow\quad\begin{cases}z=\pm\alpha\,u \text{ yields an optimal surface } S_z,\\ \pm\frac{1}{\alpha}\varsigma\ \text{ "unprojected" onto } S_z \text{ is an optimal stress.}\end{cases} \tag{1.10}$$

Of course this is a mere sketch of the construction. Although showing existence of solution ς to (OPM) is relatively straightforward, the solution u spoken of above must be intended in some relaxed sense, it is not *a priori* clear whether it is continuous.

The foregoing outline of the construction of the form-structure is based on intuition drawn from theories of optimal funiculars and arch-grids – the proof of optimality is yet to be made. The plan, however, is more ambitious than showing optimality just amongst the form-structures. It is to show that the form thus constructed is of minimum volume amongst all the 3D structures provided that the load is vertically transmissible. Such a result would be an explanation behind the many numerical simulations in which the 3D structures concentrate around a single surface, see [88, 92, 107, 117].

The 3D problem of optimal design under transmissible loads – herein called the *Prager problem* to celebrate the works [143, 146] – has never been quite posed mathematically. Its formulating is the aim of Section 6.2 where for 3D domain $\Omega = \Omega\times(-h_0,h_0)$ we start with a measure-theoretic definition of the set of loads attainable through vertical transmission of f. The under-bar $\underline{(\cdot)}$ is to stress three-dimensionality of an object. If the loading data is a positive measure, namely $f\in\mathcal{M}_+(\overline{\Omega})$, then the set reads:

$$\mathscr{T}(\underline{\Omega},f) = \Big\{\underline{F} = \mathrm{e}_3\,\underline{f}\ :\ \underline{f}\in\mathcal{M}_+\big(\overline{\underline{\Omega}}\big),\ \ \underline{f}(B\times\mathbb{R}) = f(B)\quad\forall\text{ Borel set } B\subset\overline{\Omega}\Big\}.$$

The proposal of the *Plastic setting of Prager Problem* for the pure-tension variant is as follows

$$\underline{\mathfrak{Z}} := \inf_{\substack{\underline{F}\in\mathcal{M}(\overline{\Omega};\mathbb{R}^3)\\ \underline{\varsigma}\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{3\times 3}_{+})}}\left\{\int_{\overline{\underline{\Omega}}}\underline{\rho}^0(\underline{\varsigma})\ :\ \underline{F}\in\mathscr{T}(\underline{\Omega},f),\ \ -\mathrm{Div}\,\underline{\varsigma} = \underline{F}\ \ \text{in }\mathbb{R}^3\setminus\underline{\Upsilon}_0\right\} \tag{PPP}$$

where $\underline{\Upsilon}_0 = \partial\Omega\times\{0\}$ is the curve on which the structure may rest. The function $\underline{\rho}^0$ is the polar to the spectral norm hence $\underline{\rho}^0(\underline{\varsigma}) = \mathrm{Tr}\,\underline{\varsigma}$ due to positive semi-definiteness of $\underline{\varsigma}$. The Prager problem is again close to the one of Michell: aside from restricting to tension the major difference is designing the load. The *Elastic setting of Prager Problem* (EPP) that

is equivalent to (PPP) can be posed as well thanks to the effort made in Chapter 3 on the anisotropy design: it is the (FMD) formulation extended by the choice of $\underline{F} \in \mathscr{T}(\underline{\Omega}, f)$ and specified for the *Canvas-like Material Design* setting (see Example 3.28) where

$$\mathcal{K} = \mathcal{K}_{fib}, \qquad j = j_+, \qquad c = \mathrm{Tr}. \tag{1.11}$$

Thanks to the weak-* compactness of the set $\mathscr{T}(\underline{\Omega}, f)$ existence of solutions to both Prager problems (PPP), (EPP) is established in Theorem 6.8. The goal is to show that there exists a solution $(\underline{\varsigma}, \underline{F})$ for (PPP) that constitutes a form, namely for which there exists an elevation function $z \in C_0(\Omega; \mathbb{R})$ such that the pair of measures is concentrated on its graph:

$$\mathrm{spt}\,(\underline{\varsigma}, \underline{F}) \ \subset \ \overline{S}_z. \tag{1.12}$$

The method is precisely to build on the solution ς of the optimal membrane problem (OPM) together with its deflection u and to utilize the scheme in (1.10). With $z = \alpha u$ the load $\underline{F}$ can be realized through the push-forward of $\mathrm{e}_3\, f$ by $\underline{z} = \underline{z}(x) = \big(x, z(x)\big)$. The unprojection of the plane stress $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{2\times 2}_+)$ to the 3D stress $\underline{\varsigma} \in \mathcal{M}(\overline{\underline{\Omega}}; \mathcal{S}^{3\times 3}_+)$ concentrated on $\overline{S}_z$ is more involved (see Section 6.3).

After the construction of the pair $(\underline{\varsigma}, \underline{F})$ the proof of optimality for (PPP) awaits. For this purpose a deep analysis of the newly derived optimal membrane problem (OPM) is essential. It is why the Chapter 5 on (OPM) precedes the one on optimal forms. Depending on the level of comprehension of the (OPM) problem a less or more general results about the optimal forms can be stated. Eventually, the desired solution $(\underline{\varsigma}, \underline{F})$ satisfying (1.12) is constructed in Theorem 6.18, although an extra assumption on regularity of solution of (OPM) is required. It is the author's belief that this result is – in terms of mechanics – the most significant one in this thesis.

The analytical work is followed by the computer method: once the numerical scheme is developed for the problem (OPM), the approximations of its solutions may be used to construct the numerical form-like solutions of the Prager problem (PPP) (see Section 6.5). This way may focus on discretizing the 2D region $\overline{\Omega}$ instead of the three-dimensional design domain $\overline{\underline{\Omega}} = \overline{\Omega} \times [-h_0, h_0]$. This gives the edge over the 3D methods known in the literature – both in terms of computational efficiency and quality of the results.

1.2.3. A new design formulation – optimal membrane problem

On one hand the optimal membrane problem (OPM) can be seen as means of constructing an optimal form. On the other, it is an independent design problem that proves to be fascinating – not only in terms of mechanics but above all from the mathematical point of view. The completely new (OPM) problem was the topic of a joint research with Guy Bouchitté, cf. the preprint [29]. The results are covered in Chapter 5 being by far the longest and the most technical in this thesis.

Mathematically the plastic variant of the optimal membrane problem (OPM) resembles the Free Material Problem (FMD) written for the physical context of stationary heat conduction, i.e. when u is scalar and the anisotropy field λ is valued in $\mathcal{S}^{2\times 2}_+$. Now the physics changes as λ, representing distribution of conductivity tensor before, becomes the

pre-stress ς, but the real modification lies in the zero-divergence constraint $\operatorname{Div}\varsigma = 0$ in Ω. It spoils the theory of (FMD) put forth in Chapter 4, in particular the link with the pair of problems $(\mathcal{P}), (\mathcal{P}^*)$.

Notwithstanding the issues pointed out above a pursuit of a pair of the type $(\mathcal{P}), (\mathcal{P}^*)$ that underlies the optimal membrane problem is worthwhile and, as it unravels, possible (see Section 5.2.3). Once the problem (OPM) gets rewritten by introducing the Lagrange multiplier w for the equilibrium equation $\operatorname{Div}\varsigma = 0$ and after additional manipulations we reach a problem:

$$\mathcal{Z} := \sup\left\{\int_{\overline{\Omega}} u\,df \;:\; (u,w) \in \mathcal{D}(\Omega;\mathbb{R}\times\mathbb{R}^2),\quad \mathcal{A}(u,w) \preceq \mathrm{I} \;\text{ in } \overline{\Omega}\right\} \tag{$\mathscr{P}$}$$

where a differential operator is introduced:

$$\mathcal{A}(u,w) := \tfrac{1}{2}\,\nabla u \otimes \nabla u + e(w).$$

Due to non-linearity of the operator $\mathcal{A}$ a relation with the class of problems $(\mathcal{P}), (\mathcal{P}^*)$ is not obvious. Nonetheless the set of (u,w) that are admissible for $(\mathscr{P})$ turns out to be convex. Ultimately, there is a closed gauge $\varrho : \mathbb{R}^2 \times \mathcal{S}^{2\times 2} \to \mathbb{R}_+$ that renders the equivalence:

$$\varrho\big(\nabla u, e(w)\big) \le 1 \quad \text{in } \overline{\Omega} \qquad \Leftrightarrow \qquad \mathcal{A}(u,w) \preceq \mathrm{I} \quad \text{in } \overline{\Omega}. \tag{1.13}$$

Accordingly, $(\mathscr{P})$ can be reformulated as a particular case of $(\mathcal{P})$ after all: indeed, it is enough to take $U = \mathbb{R}\times\mathbb{R}^2$, $V = \mathbb{R}^2 \times \mathcal{S}^{2\times 2}$, $A(u,w) = \big(\nabla u, e(w)\big)$, $\rho = \varrho$, $\mathscr{F}(u,w) = \int_{\overline{\Omega}} u\,df$. Automatically, specifying $(\mathcal{P}^*)$ in this framework yields the dual to $(\mathscr{P})$:

$$\mathcal{Z} = \inf\left\{\int_{\overline{\Omega}} \varrho^0(\vartheta,\varsigma) \;:\; (\vartheta,\varsigma) \in \mathcal{M}(\overline{\Omega};\mathbb{R}^2\times\mathcal{S}^{2\times 2}_+),\; -\operatorname{div}\vartheta = f,\;\; \operatorname{Div}\varsigma = 0 \;\text{ in } \Omega\right\} \tag{$\mathscr{P}^*$}$$

For $(q,\sigma) \in \mathbb{R}^2 \times \mathcal{S}^{2\times 2}_+$ the formula for the polar function is $\varrho^0(q,\sigma) = \operatorname{Tr}\sigma + \frac{1}{2}\langle \sigma^{-1}q, q\rangle$ where the latter term is understood as ∞ once q is outside the range of σ. The new variable ϑ models the field of *transverse force* that emerges under the deflection of the membrane. Solution (ϑ,ς) of $(\mathscr{P}^*)$ exists (see Proposition 5.14) and then, upon multiplying by a suitable constant, ς solves (OPM) (see Proposition 5.20).

The formulation of Optimal Plastic Membrane (OPM) is far from being justified mechanically. In some aspect it reminds of Michell problem (1.5) where the stresses are controlled as well. On the other hand it considerably differs by the minimized functional $\mathcal{C}(\varsigma)$ that involves deflection u. The Michell problem can be recast through a special formulation of an elastic design problem, see Section 4.7 or [5, 15, 46, 34]. An analogous goal is set for the optimal membrane problem, but for that purpose a new model of membrane must be put forth (see Section 5.2). The clue may be found in the spontaneously emerging operator $\mathcal{A}$ which is known from the geometrically non-linear von Kármán plate model: if, next to deflection u, the function w is the in-plane displacement of the central plane then $\frac{1}{2}\,\nabla u \otimes \nabla u + e(w)$ is the in-plane strain. If membrane is treated as a very thin plate then the bending effects can be omitted and $\mathcal{A}(u,w)$ is the strain measurement that we are looking for. To avoid issues of non-convexity known from the von Kármán theory the anisotropy framework established in Chapter 3 comes to our aid once again. Upon choosing the energy function of type j_+ that dismisses compressive stresses – being

justified especially if the membrane is made of canvas – the proposal of a new membrane model goes through the definition of the membrane's compliance:

$$\mathscr{C}(\lambda) := \sup\left\{\int_{\overline{\Omega}} u\,df - \int_{\overline{\Omega}} j_+\big(\lambda, \mathcal{A}(u,w)\big) \;:\; (u,w) \in \mathcal{D}\big(\Omega; \mathbb{R}\times\mathbb{R}^2\big)\right\}$$

where $\lambda \in \mathcal{M}\big(\overline{\Omega}; \mathscr{L}_+(\mathcal{S}^{2\times 2})\big)$ is a distribution of Hooke tensor. The *Optimal Elastic Membrane* problem can be readily formulated in the spirit of the Free Material Design:

$$\mathscr{C}_{\min} := \min\left\{\mathscr{C}(\lambda) \;:\; \lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K}),\ \int_{\overline{\Omega}} c(\lambda) \le C_0\right\} \tag{OEM}$$

If the Canvas-like Material Design setting (1.11) is chosen, in particular $\mathcal{K} = \mathcal{K}_{fib}$, then the precise link between (OEM) and the pair $(\mathscr{P}), (\mathscr{P}^*)$ can be made (see Theorem 5.17). What is more, the elastic design problem (OEM) becomes an independent problem if other triples $\mathcal{K}, j_+, c$ are considered.

The first great challenge in analysis of the pair of problems $(\mathscr{P}), (\mathscr{P}^*)$ is relaxation of the condition on the smoothness of (u,w) in $(\mathscr{P})$ so that existence of solution can be claimed (see Section 5.3). The route leads through a two-point condition that proves to be equivalent to either of the conditions in (1.13):

$$\frac{1}{2}\,|u(y)-u(x)|^2 + \langle w(y)-w(x), y-x\rangle \le |y-x|^2 \qquad \forall\,(x,y) \in \overline{\Omega}\times\overline{\Omega}. \tag{1.14}$$

The key observation follows: for any admissible smooth pair (u,w) the function

$$v = \mathrm{id} - w$$

is a maximal monotone map in the sense of [3], more precisely the uni-valued multifunction $\mathbf{v} = \mathbf{v}(x) = \{v(x)\} = \{x - w(x)\}$ is an element of the set

$$\mathbf{V}_\Omega := \Big\{\mathbf{v} : \mathbb{R}^2 \to 2^{\mathbb{R}^2} \;:\; \mathbf{v} \text{ maximal monotone},\ \mathbf{v} = \mathrm{id} \quad \text{in } \mathbb{R}^2 \setminus \overline{\Omega}\Big\}.$$

By unlocking the toolbox on maximal monotone maps developed in [3] the compactness result can be established (see Theorem 5.25). The closure of the set of admissible functions (u,w) in $C(\overline{\Omega})\times L^1(\Omega;\mathbb{R}^2)$, denoted by $\overline{\mathcal{X}}_1$, is a bounded subset of

$$\Big((C^{0,\frac{1}{2}} \cap W^{1,2})(\Omega;\mathbb{R})\Big) \times \Big((BV \cap L^\infty)(\Omega;\mathbb{R}^2)\Big)$$

and thus $\overline{\mathcal{X}}_1$ is compact, which guarantees existence of solution of the relaxed problem $(\overline{\mathscr{P}})$. Further exploitation of the theory of maximal monotone maps leads to the characterization of the set $\overline{\mathcal{X}}_1$ in Theorem 5.32, being one of the most technical results in this thesis.

The next challenge consists in deriving the optimality conditions for the pair $(\overline{\mathscr{P}}), (\mathscr{P}^*)$ (see Section 5.4). This part is absolutely essential in the view of application to constructing optimal forms. The issue is that the theory of μ-tangential calculus does not operate on arbitrary functions (u,w) in the set $\overline{\mathcal{X}}_1$. Unfortunately, at this stage of research, an extra assumption has to be made: the candidates (u,w) for solutions of $(\overline{\mathscr{P}})$ must be Lipschitz continuous. Then, apart from the admissibility conditions for (u,w) and (ϑ,ς), the system of optimality conditions reads (see Theorem 5.41):

$$\begin{cases} \big\langle \frac{1}{2}\,\nabla_\mu u \otimes \nabla_\mu u + e_\mu(w), \sigma\big\rangle = \operatorname{Tr}\sigma \quad \mu\text{-a.e.},\\ q = \sigma\,\nabla_\mu u \quad \mu\text{-a.e.}, \end{cases}$$

where a decomposition $(\vartheta, \varsigma) = (q\mu, \sigma\mu)$ is made. The optimality condition are then employed to analytically solve the radial-symmetric class of optimal membrane problem.

The efficient method of tackling the Free Material Design of heat conductor leads through optimal transport problem (OTP) of Monge and Kantorovich (see Section 4.8). The similarity between this variant of (FMD) and the problem (OPM) sparks a hope of connecting (OPM) to (OTP) as well, although the new constraint $\operatorname{Div}\varsigma = 0$ stands in the way. Ultimately, again upon the change of variable $\mathbf{v}(x) = \{x - w(x)\}$, the problem $(\overline{\mathscr{P}})$ admits a *geometric reformulation* (see Section 5.6):

$$\mathcal{Z} = \sup_{(u,\mathbf{v})\in C_0(\Omega)\times\mathbf{V}_\Omega} \left\{ \int_{\overline{\Omega}} u\, df \;:\; u(y) - u(x) \le c_{\mathbf{v}}(x,y) \quad \forall\,(x,y)\in\overline{\Omega}\times\overline{\Omega} \right\} \qquad (\mathscr{P}_{\mathrm{geo}})$$

where $c_{\mathbf{v}}$ is a pseudo-metric that is well-behaved with respect to topology on $\mathbf{V}_\Omega$ (see Theorem 5.82). When $\mathbf{v} = \{v\}$ is univalued and smooth $c_{\mathbf{v}}$ is a geodesic distance with respect to the Riemannian metric generated by the metric tensor $\sqrt{2}\,e(v)$. The supremum with respect to u in $(\mathscr{P}_{\mathrm{geo}})$ can be rewritten using the famous Kantorovich-Rubinsten theorem (see [166]) thus recovering the Wasserstein distance with respect to the cost $c_{\mathbf{v}}$:

$$W_{c_{\mathbf{v}}}^{\partial\Omega}(f_+, f_-) := \min\left\{ \int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}(x,y)\,\gamma(dxdy) \;:\; \gamma\in\Gamma^{\partial\Omega}(f_+,f_-) \right\}.$$

The completely unexpected result is that the number $\mathcal{Z} = \sup(\mathscr{P}) = \min(\mathscr{P}^*)$ that characterizes the optimal membrane problem can be recovered as the Monge-Kantorovich cost that is maximal with respect to cost functions induced by maximal monotone maps (see Theorem 5.86):

$$\mathcal{Z} = \max\left\{ W_{c_{\mathbf{v}}}^{\partial\Omega}(f_+,f_-) \;:\; \mathbf{v}\in\mathbf{V}_\Omega \right\}.$$

Despite the effort made in Section 5.6, the connection between the pre-stress ς solving (OPM) and the optimal transportation plan γ problem for "the worst metric $c_{\mathbf{v}}$" remains mysterious for the time being (see Section 5.7.4). Notwithstanding this, the discovered connection is, to the best knowledge of the present author, the first between an optimal design problem in elasticity and the classical Monge-Kantorovich problem.

Finally, the two-point condition (1.14) opens up a possibility for an alternative duality scheme: between $(\mathscr{P})$ an its another dual $(\widetilde{\mathscr{P}^*})$, whilst the proof of the zero duality gap is complex and technical (see Theorem 5.50). In the problem $(\widetilde{\mathscr{P}^*})$ we are looking for pairing measures $(\pi, \mathscr{S}) \in \mathcal{M}(\overline{\Omega}\times\overline{\Omega}; \mathbb{R}\times\mathbb{R}_+)$ that represent membrane which decomposes into a truss-like system of strings. Sadly, the problem $(\widetilde{\mathscr{P}^*})$ does not have a solution in general, although there are strong implications towards existence when f is finitely supported (see Conjecture 5.106). More importantly, the pair of problems $(\mathscr{P}), (\widetilde{\mathscr{P}^*})$ serves as a springboard for the numerical simulations of optimal membranes by the ground structure method (see Sections 5.5.4 - 5.5.6). The quadratic two-point constraint (1.14) rules out the linear programming formulation known from the classical ground structure method for trusses [91, 172, 158]. Instead, an equally efficient conic quadratic programming formulation can be applied. The 2D numerical method thus developed translates to the 3D problem of optimal form-finding where it serves as a highly efficient tool for constructing numerical grid-shell approximations of optimal forms.

Chapter 2

Mathematical preliminaries

This chapter is to gather the main mathematical tools used throughout this thesis. For the most part literature references are given, although some of the theorems are proved either for illustrative purpose or to customize. Considering the interdisciplinary nature of the thesis it was deemed worthwhile to complement the results with a short review of the corresponding topics.

2.1. Convex analysis and calculus of variations

Let $\overline{\mathbb{R}} = [-\infty, \infty]$ stand for the set of real numbers extended by $\{-\infty, \infty\}$. For X being any linear space let us consider an extended-real convex function (or a *convex functional* if X is infinite dimensional) $f : X \to \overline{\mathbb{R}}$. We say that f is *proper* if it nowhere takes the value $-\infty$ and it is not identically equal to ∞. The *effective domain* of f is the set $\operatorname{dom}(f) := \big\{x \in X \,:\, f(x) < \infty\big\}$.

For a convex set $A \subset X$ by a *convex minimization problem* (P) or a *convex program* (P) we shall call

$$\mathscr{Z} = \inf\Big\{f(x) \,:\, x \in A\Big\} \tag{P}$$

The function f will be called the *objective function* of (P), while the number $\mathscr{Z}$ will be referred to as the *objective value* or the *value function* of problem (P). We will say that A is the *admissible set* of (P) and that x is admissible for (P) when $x \in A$. We will denote $\inf \mathrm{P} := \mathscr{Z}$.

If A is an empty set then we agree that $\mathscr{Z} = \infty$ by convention. This makes it possible to reformulate (P) by using the the convex *indicator function*:

$$\mathscr{Z} = \inf\Big\{f(x) + \chi_A(x) \,:\, x \in X\Big\} \tag{2.1}$$

where for any set $A \subset X$ (not necessarily convex)

$$\chi_A(x) := \begin{cases} 0 & \text{if } x \in A, \\ \infty & \text{if } x \in X \setminus A. \end{cases}$$

Note that if f is defined only on A then it may be extended by ∞ to whole X so that (2.1) is meaningful.

Definition 2.1. For a convex function $f : X \to (-\infty, \infty]$ we say that problem (P) *admits a solution* (or that *minimum is achieved for* (P)) if the two conditions are met:

(i) $\inf \mathrm{P} < \infty$,
(ii) there exists $\bar{x} \in A$ such that $f(\bar{x}) = \inf \mathrm{P}$.

Then, $\bar{x}$ is termed a *solution* of the problem (P) and we write

$$f(\bar{x}) = \min \mathrm{P}.$$

Remark 2.2. In many text the condition (i) in Definition 2.1 is omitted. We see that (i) is equivalent to conditions: f is proper, A is non-empty and $\mathrm{dom}(f) \cap A \neq \varnothing$. This way we rule out the trivial scenario when $f(x) = \infty = \inf \mathrm{P}$ for every $x \in A$. It also allows to treat the problem (2.1) as equivalent to the original (P): otherwise for $A = \varnothing$ each $x \in X$ would be a solution of (2.1) while clearly (P) would admit none.

We put forth one of the two main tools that will be used in this thesis for showing existence of solutions, i.e. the well established Direct Method of Calculus of Variations:

Theorem 2.3 (**Direct Method of Calculus of Variations**). *Let there exist a topology on X such that $f : X \to (-\infty, \infty]$ is lower semi-continuous and A is a non-empty compact set. Then solution of problem* (P) *exists provided that* $\mathrm{dom}(f) \cap A \neq \varnothing$.

Proof. See for instance [71]. □

Remark 2.4. Validity of Theorem 2.3 *a priori* does not rely on convexity of either f or A. However, for infinite dimensional space X in practice it is difficult to provide the required topology without the convexity assumptions.

Theorem 2.5 (**Ky Fan's min-max theorem**). *For X being a locally convex topological linear space and for Y being any linear space let $A \subset X$, $B \subset Y$ be non-empty convex sets. For a real function $\mathcal{L} : A \times B \to \mathbb{R}$ assume the following:*

(i) for each $y \in B$ the function $\mathcal{L}(\,\cdot\,, y) : A \to \mathbb{R}$ is convex and lower semi-continuous,
(ii) for each $x \in A$ the function $\mathcal{L}(x, \cdot\,) : B \to \mathbb{R}$ is concave,
(iii) A is compact.

Then the following inf-sup equality holds:

$$\inf_{x \in A} \sup_{y \in B} \mathcal{L}(x, y) = \sup_{y \in B} \min_{x \in A} \mathcal{L}(x, y),$$

where the minimum on the right hand side is achieved.

If in addition Y is locally convex topological linear space and:

(iv) for each $x \in A$ the function $\mathcal{L}(x, \cdot\,) : B \to \mathbb{R}$ is upper semi-continuous,
(v) B is compact,

then above $\inf, \sup$ *may be replaced by* $\min, \max$ *and $\mathcal{L}$ admits a saddle point $(\bar{x}, \bar{y}) \in A \times B$, i.e.*

$$\mathcal{L}(\bar{x}, y) \leq \mathcal{L}(\bar{x}, \bar{y}) \leq \mathcal{L}(x, \bar{y}) \qquad \forall\, (x, y) \in A \times B.$$

Proof. The result is a variant of Ky Fan's theorem, for the proof see [171, Theorem 2.10.2, p. 144]. □

Below we gather some useful facts on topological properties of convex functions:

Proposition 2.6 (**Continuity of convex functions**). *For X being a locally convex topological linear space let $f : X \to \overline{\mathbb{R}}$ be a proper convex function. Assume that there exists an open set $\mathscr{U} \subset X$ such that $\sup_{x\in\mathscr{U}} f(x) < \infty$.*

Then f is continuous in $\operatorname{int}\big(\operatorname{dom}(f)\big)$, i.e. in the interior of its effective domain.

Proof. See [80, Chapter I, Proposition 2.5] □

Corollary 2.7. *For a finite dimensional space X every convex real function $f : X \to \mathbb{R}$ is continuous on X.*

Proof. See [80, Chapter I, Corollary 2.3] □

Proposition 2.8. *For X being a locally convex topological linear space let $\mathscr{G}$ be any family (it can be uncountable) whose elements $g : X \to \overline{\mathbb{R}}$ are convex lower semi-continuous functions on X. Then, the function $f : X \to \overline{\mathbb{R}}$ being the point-wise supremum of this family, i.e.*

$$f(x) := \sup_{g\in\mathscr{G}} g(x),$$

is convex and lower semi-continuous as well.

Proof. See [80, Chapter I, Sections 2.1, 2.2]. □

Let us now move on to discuss elements of duality theory. For a locally convex linear topological space X we denote by X^* its dual space, i.e. the space of linear continuous functionals over X. For $(x, x^*) \in X \times X^*$ we introduce the duality pairing: $\langle x, x^*\rangle := x^*(x)$.

Let us now take any function $f : X \to \overline{\mathbb{R}}$, i.e. it does not have to be convex or lower-semicontinuous. By *convex conjugate* and *convex biconjugate* of function f with respect to the duality X, X^* we call the functions $f^* : X^* \to \overline{\mathbb{R}}$ and, respectively, $f^{**} : X \to \overline{\mathbb{R}}$ defined by:

$$f^*(x^*) := \sup_{x\in X} \Big\{\langle x, x^*\rangle - f(x)\Big\}, \qquad f^{**}(x) := \sup_{x^*\in X^*} \Big\{\langle x, x^*\rangle - f^*(x^*)\Big\}.$$

The function f^* is also often referred to as *Lagandre-Fenchel transform* (note that in [80] it is called the polar function – herein this term will be reserved for other notion, see below).

Proposition 2.9. *For any function $f : X \to \overline{\mathbb{R}}$ the functions f^* and f^{**} are convex and lower semi-continuous with respect to, respectively, weak-* topology on X^* and weak topology on X.*

Proof. For every $x \in X$ the mapping $X^* \ni x^* \mapsto \langle x, x^*\rangle$ is continuous for the weak-* topology by the very definition of the latter. Therefore the set $\mathscr{G} = \big\{\langle x, \cdot\,\rangle - f(x) : x \in X\big\}$ is a family of affine (and thus convex) continuous functions on X^* equipped with the weak-* topology, hence the claim for f^* by Proposition 2.8. The proof for f^{**} is similar. □

Remark 2.10. In a locally convex linear topological space a convex function $f : X \to \overline{\mathbb{R}}$ is weakly lower semi-continuous if and only if it is lower semi-continuous for the original topology on X, see [80, Chapter I, Corollary 2.2].

An important convex conjugate is the one of the indicator function χ_A for a subset $A \subset X$: it attains the formula

$$\chi_A^*(x^*) = \sup_{x \in X} \Big\{ \langle x, x^* \rangle \ : \ x \in A \Big\}.$$

The function $\chi_A^* : X^* \to \overline{\mathbb{R}}$ is called *the support function of a set* A.

It is well known that a function $f : X \to (-\infty, \infty]$ is convex and weakly lower semi-continuous if and only if $f = f^{**}$, see [80, Chapter I, Proposition 4.1]. In due course of this thesis we will require a more local variant of this statement:

Lemma 2.11 (**Criterion for the zero duality gap**). *For a locally convex linear topological space X let $h : X \to (-\infty, \infty]$ be a convex function such that $h(0) < +\infty$ and h is lower semi-continuous at 0. Then it holds that $h(0) = h^{**}(0) = -\inf_{x^* \in X^*} h^*(x^*)$.*

If in addition h is continuous at 0, then h^* reaches its minimum on X^*.

Proof. See [80, Chapter III, Propositions 2.1, 2.2] □

The lemma above will be used directly in Chapter 5, but it also can be utilized to derive a very systematic tool for dualizing convex (or concave) problems. This classic duality result is stated below and it will be employed on multiple occasions throughout the thesis. For locally convex linear topological spaces X, Y and their respective duals X^*, Y^* we consider a linear continuous operator $\Lambda : X \to Y$, while the *adjoint* $\Lambda^* : Y^* \to X^*$ is the uniquely defined linear continuous operator that for each pair $x \in X$, $y^* \in Y^*$ satisfies $\langle x, \Lambda^* y^* \rangle = \langle \Lambda x, y^* \rangle$.

Theorem 2.12 (**Classical duality**). *For X, Y and Λ as above let $\Phi : X \to \overline{\mathbb{R}}$, $\Psi : Y \to \overline{\mathbb{R}}$ be convex lower semi-continuous functionals while by $\Phi^* : X^* \to \overline{\mathbb{R}}$, $\Psi^* : Y^* \to \overline{\mathbb{R}}$ denote their convex conjugates.*

If there exists $x_0 \in X$ such that $\Phi(x_0) < +\infty$ and Ψ is continuous at Λx_0 then

$$\mathscr{Z} := \sup_{x \in X} \Big\{ -\Phi(x) - \Psi(\Lambda x) \Big\} = \inf_{y^* \in Y^*} \Big\{ \Phi^*(-\Lambda^* y^*) + \Psi^*(y^*) \Big\} \tag{2.2}$$

and the minimum on the right hand side is achieved provided that $\mathscr{Z} < \infty$.

Proof. See Theorem 4.1, equation (4.18) and condition (4.21) in Chapter III of [80]. □

Remark 2.13. In Lemma 2.11 and in Theorem 2.12 the lower semi-continuity and continuity prerequisites are intended for the original topologies of the spaces X, Y, not the weak ones. This makes them easier to satisfy. For the lower semi-continuity this fact is related to Remark 2.10, whilst it is a more delicate matter for continuity, cf. [80] for details.

A proper convex function $f : X \to \overline{\mathbb{R}}$ is called *subdifferentiable* at $x \in \mathrm{dom}(f)$ if there exists a continuous functional $x^* \in X^*$ such that for some $a \in \mathbb{R}$ the affine function $\langle \cdot, x^* \rangle + a$ minorizes f and moreover $f(x) = \langle x, x^* \rangle + a$, i.e. the two functions agree at x. Every such x^* is called a *subgradient* of f at x and the set of all subgradients is called the *subdifferential* at x and it is denoted by $\partial f(x)$.

Proposition 2.14. *For every x the set $\partial f(x)$ is a (possibly empty) convex and weakly-* closed subset of X^*. Moreover, the following conditions are equivalent for any pair $(x, x^*) \in X \times X^*$:*

(i) $x^* \in \partial f(x)$,
(ii) $\langle x, x^* \rangle = f(x) + f^*(x^*)$,
(iii) $\langle x, x^* \rangle \geq f(x) + f^*(x^*)$.

Proof. See [80, Chapter I, Proposition 5.1 and Corollary 5.1]. □

Remark 2.15. The inequality that is opposite to (iii) holds always by definition of f^* and it is known as the *Fenchel's inequality*.

Let us wrap up the review on convex analysis by considering the class of functions $g : X \to [0, \infty]$ that are non-negative, convex, lower semi-continuous and positively one-homogeneous, i.e. $g(tx) = t\, g(x)$ for any $x \in X$, $t \geq 0$. For such functions we define the polar $g^0 : X^* \to \overline{\mathbb{R}}$ via the formula

$$g^0(x^*) := \sup_{x \in X} \Big\{ \langle x, x^* \rangle \,:\, g(x) \leq 1 \Big\}.$$

Considering Proposition 2.8 it is easy to show that g^0 is also non-negative convex positively one-homogeneous and lower semi-continuous (with respect to the weak-* topology). Convex and one-homogeneous function may be constructed as *gauges* of convex sets $\mathrm{C} \subset X$:

$$g_{\mathrm{C}}(x) := \inf_{t > 0} \left\{ t \,:\, \frac{1}{t}\, x \in \mathrm{C} \right\}.$$

When X is infinite dimensional g_{C} is more often known as the *Minkowski functional of* C. The function g_{C} is lower semi-continuous if and only if C is closed. If this is the case g_{C} is termed a *closed gauge*. It is well known (see [140, Section 15]) that for any non-negative convex lower semi-continuous positively one-homogeneous function g there exists a unique closed convex set C that contains the origin for which $g = g_{\mathrm{C}}$. This one-to-one correspondence justifies calling each such function g a closed gauge.

It is well known (and in fact easy to show) that the polar of a gauge function of a closed set C is equal to the support function of the latter, i.e.

$$g_{\mathrm{C}}^0 \;=\; \chi_{\mathrm{C}}^*. \tag{2.3}$$

2.2. Measure theory and convex integral functionals

We will now review some basic elements of measure theory; to large extent we shall follow the book [8] where the reader is referred for details. For a set X we choose a family of its subsets Σ that forms a σ-algebra. The pair (X, Σ) is then called a *measurable space*. We say that a function $\mu : \Sigma \to [0, \infty]$ is a *non-negative measure* if it is σ-additive (i.e. countably additive) and $\mu(\varnothing) = 0$. The measure μ is called *finite* if $\mu(X) < \infty$. For a metric space Y a function $f : X \to Y$ is called Σ-measurable if for any open set $\mathscr{U} \subset Y$ the preimage $f^{-1}(\mathscr{U})$ belongs to Σ.

Further we shall assume that X is a locally compact and separable metric space. The topological structure on X offers a natural choice of the σ-algebra Σ. The *Borel σ-algebra*, denoted by $\mathscr{B}(X)$, is the smallest σ-algebra containing all the open subsets of X. Any measure μ on $\mathscr{B}(X)$ is called a *Borel measure*. A function $f : X \to Y$ is called *Borel measurable* if it is $\mathscr{B}(X)$-measurable. A Borel measure μ that is finite on compact subsets of X is called a *Radon measure*. The set of non-negative Radon measures on X will be denoted by $\mathcal{M}_+(X)$. In addition, $\mathcal{M}_+^b(X)$ will be its subset consisting of measures that are finite. A finite Radon measures $\mu \in \mathcal{M}_+^b(X)$ will be called a *probability measure* if $\mu(X) = 1$, while the set of all probability measures on X will be denoted by $\mathrm{P}(X)$.

Example 2.16. For $X = \mathbb{R}^d$ the most basic example of a non-negative Radon measure is the one of Lebesgue measure $\mathcal{L}^d$. Note that $\mathcal{L}^d(\mathbb{R}^d) = \infty$, although $\mathcal{L}^d\big(B(0,R)\big) = \omega_d R^d < \infty$ for any $R > 0$, where ω_d is the volume of a d-dimensional unit ball. Therefore $\mathcal{L}^d$ is an element of $\mathcal{M}_+(\mathbb{R}^d)$ but not of $\mathcal{M}_+^b(\mathbb{R}^d)$. An example of a finite Radon measure would be the Dirac delta measure $\delta_{x_0} \in \mathcal{M}_+^b(X)$ that for $x_0 \in X$ yields for a Borel set B: $\delta_{x_0}(B) = 1$ if $x_0 \in B$ and $\delta_{x_0}(B) = 0$ if $x_0 \notin B$. △

The notion of Radon measure extends to set functions $\vartheta : \mathscr{B}(X) \to V$ where V is a Euclidean space, endowed with Euclidean norm $|\cdot|$. We say that ϑ is V-valued or vector valued. To such measure ϑ we may assign its *variation measure* – it is the smallest non-negative Borel measure that satisfies $|\vartheta(B)| \leq |\vartheta|(B)$ for any Borel set B, see [8, Definition 1.4]. In order that the σ-additivity is met for ϑ it is necessary that $|\vartheta|(X) < \infty$. This set of such measures ϑ will be denoted by $\mathcal{M}^b(X;V)$ and it is natural to call it *a space of finite V-valued Radon measures*. The so called *total variation* $\|\vartheta\|_{\mathcal{M},X} := |\vartheta|(X)$ is a norm on the linear space $\mathcal{M}^b(X;V)$. Once X is compact we shall write $\mathcal{M}(X;V)$ instead.

If $V = \mathbb{R}$ the elements $\vartheta \in \mathcal{M}^b(X;\mathbb{R})$ are called *signed measures* and we may then define measures $\vartheta_+, \vartheta_- \in \mathcal{M}_+^b(X)$ via formulas

$$\vartheta_+ := \frac{1}{2}\big(|\vartheta| + \vartheta\big), \qquad \vartheta_- := \frac{1}{2}\big(|\vartheta| - \vartheta\big) \qquad \text{where} \quad \vartheta \in \mathcal{M}^b(X;\mathbb{R}) \tag{2.4}$$

while of course $\vartheta = \vartheta_+ - \vartheta_-$. This decomposition is known as the *Jordan decomposition* of a signed measure.

Let us now turn to the notion of Lebsegue integral with respect to $\mu \in \mathcal{M}_+(X)$. To define $\int_X f\, d\mu$ the function $f : X \to V$ must enjoy some measurability properties. Requiring that f is measurable for the Borel σ-algebra is too strong – often f is defined up to the set of zero measure, that may be not Borel. A set $N \subset X$ is called *μ-negligible* if there exists a Borel set $B \supset N$ such that $\mu(B) = 0$. Accordingly, a *μ-completion* $\mathscr{B}_\mu(X)$ of $\mathscr{B}(X)$ (or in general Σ_μ of Σ on which μ is initially defined) is the σ-algebra consisting of sets of the form $A = B \cup N$, where B is Borel and N is μ-negligible. A Radon measure μ naturally extends to $\mathscr{B}_\mu(X)$: we put $\mu(A) = \mu(B \cup N) := \mu(B)$. We will say that a property $P = P(x)$ holds μ-a.e. in a μ-measurable set $A \subset X$ if $P(x)$ is true for any $x \in A \setminus N$ where N is a μ-negligible set.

Readily, a function f is called *μ-measurable* if it is measurable with respect to $\mathscr{B}_\mu(X)$ (Σ_μ in general). For every such non-negative function the integral $\int_A f\,d\mu$ is well defined for every *μ-measurable set* A, i.e. for every $A \in \mathscr{B}_\mu(X)$. If $f : X \to V$ then the integral (valued in V) is meaningful provided that f is an element of the *Lebesgue space* $L^1_\mu(X;V)$ where

$$L^p_\mu(X;V) := \left\{ f : X \to V \,:\, f \text{ is } \mu\text{-measurable}, \quad \int_X |f|^p\,d\mu < \infty \right\}, \qquad p \in [1,\infty].$$

Analogous facts hold true for $\vartheta \in \mathcal{M}^b(X;V)$ when the completion is defined for $|\vartheta|$. The integral then makes sense if: (1) $f \in L^1_{|\vartheta|}(X;\mathbb{R})$ and then $\int_X f\,d\vartheta \in V$; (2) $f \in L^1_{|\vartheta|}(X;V)$ and then $\int_X f\,d\vartheta =: \int_X \langle f, \vartheta\rangle \in \mathbb{R}$ where $\langle\,\cdot\,,\,\cdot\,\rangle$ is the inner product in V.

We say that a vector-valued Radon measure $\vartheta \in \mathcal{M}^b(X;V)$ is *absolutely continuous* with respect to a non-negative Radon measure $\mu \in \mathcal{M}_+(X)$ – which we write $\vartheta \ll \mu$ – if for any Borel set B the fact that $\mu(B) = 0$ implies $|\vartheta|(B) = 0$. We say that two Radon measures $\vartheta_1, \vartheta_2 \in \mathcal{M}^b(X;V)$ are *mutually singular* – which we write $\vartheta_1 \perp \vartheta_2$ – if there exists a Borel set $B_1 \subset X$ such that $|\vartheta_1|(B_1) = |\vartheta_1|(X)$ and $|\vartheta_2|(X \setminus B_1) = |\vartheta_2|(X)$. Radon-Nikodym theorem states that for any $\vartheta \in \mathcal{M}^b(X;V)$, $\mu \in \mathcal{M}_+(X)$ there exists a unique pair of Radon measures $\vartheta_a, \vartheta_s \in \mathcal{M}^b(X;V)$ and a function $f \in L^1_\mu(X;V)$ such that

$$\vartheta = \vartheta_a + \vartheta_s, \qquad \vartheta_a \ll \mu, \qquad \vartheta_s \perp \mu, \qquad \vartheta_a = f\mu$$

where by the last equality we mean that $\vartheta_a(B) = \int_B f\,d\mu$ for any Borel set B. The function f is called the *Radon-Nikodym derivative* of ϑ with respect to μ. The derivative may be constructed directly; we denote it by

$$f = \frac{d\vartheta}{d\mu}.$$

If, in particular, $\vartheta \ll \mu$ then $\vartheta = \frac{d\vartheta}{d\mu}\,\mu$.

For a Borel measure μ and a μ-measurable set $C \subset X$ (defined as for a Radon measure) by restriction $\mu \llcorner C$ we understand a set function that for any μ-measurable set $A \subset X$ yields $(\mu \llcorner C)(A) := \mu(A \cap C)$. If $\mu(C) < \infty$ then the restriction is a finite Radon measure on X, namely $\mu \llcorner C \in \mathcal{M}^b_+(X)$.

Example 2.17. For $X = \mathbb{R}^d$ and $k \in [0,d]$ by $\mathcal{H}^k$ we denote the Hausdorff measure. For natural k and a set S_k being k-dimensional manifolds of e.g. class C^1 the number $\mathcal{H}^k(S_k)$ returns the geometrical measurement of S_k: the length of a curve ($k = 1$) the area of a surface ($k = 2$), the volume of a solid ($k = 3$). The measure $\mathcal{H}^k$ is Borel but it is not Radon unless $k = d$, e.g. $\mathcal{H}^1(A) = \infty$ for any subset of $A \subset \mathbb{R}^2$ of non-zero Lebesgue measure. Accordingly, a restriction $\mathcal{H}^k \llcorner C$ is a Radon measure provided that C is k-rectifiable, e.g. $\mathcal{H}^1 \llcorner C$ for a curve C of finite length. △

Next, we say that a measure $\vartheta \in \mathcal{M}^b(X;V)$ *concentrates* on a $|\vartheta|$-measurable set $A \subset X$ if $|\vartheta|(X \setminus A) = 0$. By the *support* $\operatorname{spt}\vartheta$ of a measure ϑ we call the smallest closed set on which the measure concentrates.

Remark 2.18. If K is a compact subset of the metric space X then it is often convenient to think of $\mathcal{M}(K;V)$ as the set of Radon measures on the whole X whose support is contained in K, i.e.

$$\mathcal{M}(K;V) \equiv \Big\{\vartheta \in \mathcal{M}^b(X;V) \,:\, \operatorname{spt}\vartheta \subset K\Big\}.$$

In the text, e.g. for a bounded domain $\Omega \subset \mathbb{R}^d$ this is how we will usually understand measures in $\mathcal{M}(\overline{\Omega};V)$.

If X and Z are two locally compact separable metric spaces, $\mu \in \mathcal{M}_+(X)$ a Radon measure on X and $p : X \to Z$ is a Borel measurable function (e.g. p may be continuous) then by *the push forward of μ by p* we understand the Borel measure on Z denoted by $p\,\#\,\mu$ such that

$$(p\,\#\,\mu)(B) := \mu\big(p^{-1}(B)\big) \qquad \forall\, B \in \mathscr{B}(Z) \tag{2.5}$$

where $p^{-1}(B)$ is the pre-image of B under p. The push-forward $p\,\#\,\mu$ is a Radon measure if μ is finite (in particular when X is compact). The change of variable formula holds true for any $f \in L^1_{p\#\mu}(Z;V)$:

$$\int_Z f\, d(p\,\#\,\mu) = \int_X f \circ p\, d\mu, \tag{2.6}$$

in particular we have $f \circ p \in L^1_\mu(X;V)$. The push-forward of a vector valued Radon measure $\vartheta \in \mathcal{M}^b(X;V)$ is defined analogously.

We turn to address the topological properties of the linear space $\mathcal{M}^b(X;V)$. Let $C(X;V)$ be the linear space of continuous functions; for any its element φ we define $\|\varphi\|_{\infty,X} := \sup_{x\in X} |\varphi(x)|$. If X is a compact space then $\|\cdot\|_{\infty,X}$ is a norm on $C(X;V)$ and the space is Banach. If X is non-compact then $\|\cdot\|_{\infty,X}$ can admit infinite values. To remedy this we introduce the linear space of continuous functions compactly supported in X, namely $C_c(X;V) := \big\{\varphi \in C(X;V) \,:\, \operatorname{spt}\varphi \Subset X\big\}$ where $\operatorname{spt}\varphi := \operatorname{cl}\big(\{x \in X : \varphi(x) \neq 0\}\big)$. Then $\big(C_c(X;V), \|\cdot\|_{\infty,X}\big)$ is a normed space but it is not complete. The completion of this space is denoted by $C_0(X;V)$ and endowed with the norm $\|\cdot\|_{\infty,X}$ it is Banach by construction.

Remark 2.19. If we put $X = \Omega$ where $\Omega \subset \mathbb{R}^d$ is a bounded open set then we can make the following identification that will be exploited throughout the text:

$$C_0(\Omega;V) \;\equiv\; \Big\{\varphi \in C(\overline{\Omega};V) \,:\, \varphi = 0 \text{ on } \partial\Omega\Big\},$$

i.e. $\varphi \in C(\Omega;V)$ belongs to $C_0(\Omega;V)$ if it has a continuous extension to $\partial\Omega$ by zero.

It is straightforward to check that for a Radon measure $\vartheta \in \mathcal{M}^b(X;V)$ any function $\varphi \in C_0(X;V)$ is an element of $L^1_{|\vartheta|}(X;V)$ therefore the linear functional $C_0(X;V) \ni \varphi \mapsto \int_X \langle\varphi, \vartheta\rangle \in \mathbb{R}$ is well defined. It is straightforward to see that it is continuous with respect to the supremum norm. The far less trivial fact is that any linear continuous functional on $C_0(X;V)$ is of this form:

Theorem 2.20 (**Riesz representation theorem**)**.** *For a locally compact separable metric space X let us consider the Banach space $Y = \big(C_0(X;V), \|\cdot\|_{\infty,X}\big)$. Then the topological*

dual Y^* is isometrically isomorphic to $\big(\mathcal{M}^b(X;V), \|\cdot\|_{\mathcal{M},X}\big)$ which is a Banach space itself. In particular for every $\vartheta \in \mathcal{M}^b(X;V)$ the linear functional $C_0(X;V) \ni \varphi \mapsto \int_X \langle \varphi, \vartheta \rangle \in \mathbb{R}$ is continuous while

$$|\vartheta|(X) = \sup\left\{ \int_X \langle \varphi, \vartheta \rangle \ : \ \varphi \in C_0(X;V),\ \|\varphi\|_{\infty,X} \leq 1 \right\}. \tag{2.7}$$

If X is compact then above $C_0(X;V)$ and $\mathcal{M}^b(X;V)$ may be replaced by $C(X;V)$ and $\mathcal{M}(X;V)$, respectively.

Proof. See [8, Theorem 1.54]. □

Choosing simplicity over precision, in the text we will identify the dual $Y^* = \big(C_0(X;V)\big)^*$ with the space $\big(\mathcal{M}^b(X;V), \|\cdot\|_{\mathcal{M},X}\big)$, i.e. every $\vartheta \in \mathcal{M}^b(X;V)$ is identified with the continuous linear functional $\int_X \langle \cdot\,, \vartheta \rangle$ it induces. This also establishes the duality pairing: $\langle \varphi, \vartheta \rangle_{(C,\mathcal{M})} := \int_X \langle \varphi, \vartheta \rangle$. We can thus speak of weak-* topology on $\mathcal{M}^b(X;V)$. This unlocks the use of Banach-Alaoglu thereom to obtain the famous compactness result:

Theorem 2.21 (**Weak-* compactness of measures**). *For any constant $C > 0$ the set*

$$\left\{ \vartheta \in \mathcal{M}^b(X;V) \ : \ |\vartheta|(X) \leq C \right\} \tag{2.8}$$

is weakly- compact. Moreover, the weak-* topology on this set is metrizable and, as a result, the set is sequentially weak-* compact as well.*

Proof. See [8, Theorem 1.59]. □

Remark 2.22. If X is not compact then one has to be careful while using the compactness result above. Take for instance a bounded domain $X = \Omega \subset \mathbb{R}^d$ and a sequence of points $x_n \in \Omega$ such that $d_{\partial\Omega}(x_n) < \frac{1}{n}$. Then the set $\{\delta_{x_n}\}_{n=1}^{\infty}$ is of the form (2.8) for $V = \mathbb{R}$, $C = 1$. According to Theorem 2.21 the sequence $\{\delta_{x_n}\}_{n=1}^{\infty}$ admits a cluster point $\vartheta \in \mathcal{M}^b(\Omega;\mathbb{R})$. Then ϑ must be the zero measure which can be explained by the fact that $\int_\Omega \varphi \, d\delta_{x_n} \to 0$ for every $\varphi \in C_0(\Omega;\mathbb{R})$, see Remark 2.19. Such setting is not suitable for the purposes of this thesis therefore we shall (almost) always work with the space $\mathcal{M}^b(X;V) = \mathcal{M}(X;V)$ for a compact set X. If we take $X = \overline{\Omega}$ then (up to extracting a subsequence) $\delta_{x_n} \overset{*}{\rightharpoonup} \delta_{\bar{x}}$ where $\bar{x}$ is some point on $\partial\Omega$.

In accordance to the remark above we shall henceforward work with a compact metric space X. Let us now turn our attention to convex functionals on $\mathcal{M}(X;V) = \mathcal{M}^b(X;V)$ that are lower semi-continuous with respect to the weak-* topology. Our particular interest lies in those such functionals that can be written by means of the Lebesgue integral.

Recalling that V is a Euclidean space we take a closed gauge $g : V \to [0,\infty]$, i.e. a non-negative, convex, lower semi-continuous, positively one-homogeneous function. Then to any Radon measure $\vartheta \in \mathcal{M}(X;V)$ we may assign a non-negative Borel measure denoted by $g(\vartheta)$ that for any Borel set $B \subset X$ yields

$$g(\vartheta)(B) := \int_B g\left(\frac{d\vartheta}{d\mu}(x)\right) \mu(dx) \tag{2.9}$$

where $\mu \in \mathcal{M}_+(X)$ is any Radon measure for which $\vartheta \ll \mu$; the function $\frac{d\vartheta}{d\mu}$ is the Radon-Nikodym derivative. Owing to the one-homogeneity the integral above does not

depend on the choice of such μ, although $\mu = |\vartheta|$ is the natural one. Since every norm $|\cdot|$ is a particular case of a closed gauge the measure $g(\vartheta)$ is a generalization of the variation measure $|\vartheta|$ (and in fact can be defined identically as the variation measure). Note that $g(\vartheta)$ is Borel, but not necessarily a Radon measure – the function g may admit infinite values, thus it can happen that $g(\vartheta)(B) = \infty$ for a Borel set $B \subset X$. Consequently, for a compact space X, $g(\vartheta) \in \mathcal{M}_+(X)$ if and only if $g(\vartheta)(X) < \infty$.

We take a step further by considering *convex integrals of linear growth*, namely functions $h : X \times V \to [0, \infty]$ that are lower semi-continuous (jointly, i.e. with respect to the product topology on $X \times V$) and such that for each $x \in X$ the function $h(x, \cdot) : V \to [0, \infty]$ is a gauge. Such integrals enjoys the following property: for any Radon measure $\mu \in \mathcal{M}_+(X)$ and a μ-measurable function $f : X \to V$ the mapping $x \mapsto h\big(x, f(x)\big)$ is μ-measurable as well. Accordingly, for any $\vartheta \in \mathcal{M}(X; V)$ we can define the integral

$$\int_X h\big(x, \vartheta(dx)\big) := \int_X h\Big(x, \frac{d\vartheta}{d\mu}(x)\Big)\mu(dx) \tag{2.10}$$

where $\mu \in \mathcal{M}_+(X)$ is again any Radon measure satisfying $\vartheta \ll \mu$. Note that the left integral is not *a priori* well defined, e.g. for h that is quadratic with respect to the second argument. The equality above therefore relies on the one-homogeneity of $h(x, \cdot)$ and it should be treated in terms of convention to which we shall refer as *the Goffman-Serrin convention*; definition (2.9) may be viewed as its particular case.

Let us consider non-empty closed convex sets C and $\mathscr{C}$ of, respectively, the finite dimensional space V and the Banach space $Y = C(X; V)$. Then their support functions χ^*_C and $\chi^*_{\mathscr{C}}$ are closed gauges on $V^* \equiv V$ and $Y^* \equiv \mathcal{M}(X; V)$ respectively. If we take a Radon measure $\vartheta \in \mathcal{M}(X; V)$ then $\chi^*_{\mathscr{C}}(\vartheta)$ is an element of $[0, \infty]$. On the other hand, according to convention (2.9), we may define $\chi^*_{\mathrm{C}}(\vartheta)$ which is a Borel measure. Bearing this delicate distinction in mind let us address the celebrated result by Rockafellar (see Section 2.3 below for a short review on multifunctions):

Theorem 2.23 (**Rockafellar integral representation theorem**). *Let X be a compact metric space and let V be any finite dimensional space. Assume a lower semi-continuous multifunction $\Gamma : X \to 2^V$ such that $\Gamma(x)$ is a non-empty closed convex set for any $x \in X$. We define the set of continuous selections of Γ:*

$$\mathscr{C} := \Big\{\varphi \in C(X; V) \,:\, \varphi(x) \in \Gamma(x) \quad \forall\, x \in X\Big\}. \tag{2.11}$$

Then $\mathscr{C}$ is a non-empty closed convex subset of $C(X; V)$. The function $h : X \times V \to \overline{\mathbb{R}}$ defined by

$$h(x, v) := \chi^*_{\Gamma(x)}(v)$$

is lower semi-continuous, i.e. jointly in x and v. Moreover, for any $\vartheta \in \mathcal{M}(X; V)$ we have

$$\chi^*_{\mathscr{C}}(\vartheta) = \int_X h\big(x, \vartheta(dx)\big) \tag{2.12}$$

where by definition $\chi^*_{\mathscr{C}}(\vartheta) := \sup\big\{\int_X \langle \varphi, \vartheta\rangle \,:\, \varphi \in \mathscr{C}\big\}$.

Proof. The lower semi-continuity of h is due to Theorem 2.32 while for the rest of the statement see [138, Theorem 6]. □

The above theorem will allow an effective use of duality in Theorem 2.12 in this thesis. For $Y = C(X;V)$ and $Y^* = \mathcal{M}(X;V)$ we shall often choose Ψ as the indicator function of a set $\mathscr{C}$ of the form (2.11). The formula (2.12) thus provides an integral form for the convex conjugate $\Psi^* = \chi^*_{\mathscr{C}}$. In most applications we shall put $\Gamma(x) = \mathrm{C}$ for every $x \in X$ in which case (2.12) may be considered a generalization of the integral formula for the dual norm in (2.7):

Corollary 2.24. *Let X be a compact metric space and for a finite dimensional space V choose any non-empty closed convex subset $\mathrm{C} \subset V$ (possibly unbounded). Then, for any measure $\vartheta \in \mathcal{M}(X;V)$ we have*

$$\int_X \chi^*_{\mathrm{C}}(\vartheta) = \sup\left\{ \int_X \langle \varphi, \vartheta \rangle \ : \ \varphi \in C(X;V), \quad \varphi(x) \in \mathrm{C} \ \ \forall\, x \in X \right\}.$$

where $\chi^*_{\mathrm{C}}(\vartheta)$ is defined as in (2.9).

Another corollary from Rockafellar integral representation theorem is the renown Reshetnyak lower semi-continuity result:

Corollary 2.25 (**Reshetnyak lower semi-continuity theorem**). *Let X be a compact space and by V denote any finite dimensional space. Assume an extended real function $h : X \times V \to [0, \infty]$ such that: h is lower semi-continuous jointly in both arguments and for each $x \in X$ the function $h(x, \cdot\,)$ is convex proper and positively one-homogeneous.*

Then the functional

$$\mathcal{M}(X;V) \ni \vartheta \mapsto \int_X h\big(x, \vartheta(dx)\big)$$

is weakly- lower semi-continuous.*

Proof. By virtue of [140, Corollary 13.2.1] for each $x \in X$ there exists a closed convex set $\Gamma(x)$ such that $h(x,v) = \chi^*_{\Gamma(x)}(v)$ for every $v \in V$. According to Theorem 2.32 the multifunction $x \mapsto \Gamma(x)$ is lower semi-continuous. By Theorem 2.23 (see Section 2.3 below) the integral functional is the support function $\chi^*_{\mathscr{C}}$ and the assertion follows by Proposition 2.9. □

Let us conclude this section by a construction of a Radon measure through "integration of a family of Radon measures". For a locally compact separable metric space T let us take a finite signed Radon measure $\nu \in \mathcal{M}^b(T;\mathbb{R})$. Assume that for any point $t \in T$ we have a Radon measure $\vartheta^t \in \mathcal{M}(X;V)$ were X is a compact metric space.

Definition 2.26. We shall say that the mapping $T \ni t \mapsto \vartheta^t \in \mathcal{M}^b(X;\mathbb{R})$ is ν-measurable and ν-essentially bounded if the following conditions are satisfied for any continuous function $\varphi \in C(X;V)$:

(i) the mapping $T \ni t \mapsto \int_X \langle \varphi, \vartheta^t \rangle \in \mathbb{R}$ is ν-measurable,
(ii) ν-ess $\sup_{t \in T} \left| \int_X \langle \varphi, \vartheta^t \rangle \right| < \infty$.

For the mapping $t \mapsto \vartheta^t$ that satisfies (i), (ii) above there exists a unique Radon measure $\vartheta_\nu \in \mathcal{M}(X;V)$ such that the *integration formula* holds:

$$\int_X \langle \varphi, \vartheta_\nu \rangle = \int_T \left(\int_X \langle \varphi, \vartheta^t \rangle \right) \nu(dt) \qquad \forall\, \varphi \in C(X;V). \tag{2.13}$$

Indeed, the integral on the right hand side furnishes a linear functional on $C(X;V)$ that is bounded and therefore continuous. Existence and uniqueness of ϑ_ν thus follows by Theorem 2.20: indeed, the functional must be of the form $\varphi \mapsto \langle \varphi, \vartheta_\nu \rangle_{(C,\mathcal{M})} = \int_X \langle \varphi, \vartheta_\nu \rangle$.

Corollary 2.27. *Let us take X, V, h as in Corollary 2.25. In addition let $\nu \in \mathcal{M}^b_+(T)$ (ν is non-negative) where T is a locally compact separable metric space and assume a mapping $t \mapsto \vartheta^t$ that is ν-measurable and ν-essentially bounded.*

Then, for ϑ_ν defined via (2.13)*, we have the following subadditivity property of the integral functional:*

$$\int_X h\big(x, \vartheta_\nu(dx)\big) \le \int_T \left(\int_X h\big(x, \vartheta^t(dx)\big) \right) \nu(dt). \tag{2.14}$$

Proof. Let Γ be the multifunction constructed in the proof of Corollary 2.25, in particular $h(x,v) = \chi^*_{\Gamma(x)}(v)$. Then, by non-negativity of ν, for any continuous selection φ of Γ

$$\int_X \langle \varphi, \vartheta_\nu \rangle = \int_T \left(\int_X \langle \varphi, \vartheta^t \rangle \right) \nu(dt) \le \int_T \left(\int_X h\big(x, \vartheta^t(dx)\big) \right) \nu(dt).$$

The assertion follows by taking the supremum of the left hand side with respect to $\varphi \in \mathscr{C}$ (see (2.11)) and by acknowledging (2.12). □

2.3. Multifunctions

By a multifunction Γ on a set X we understand a mapping that for every $x \in X$ returns a subset $\Gamma(x)$ of some set Y, in other words $\Gamma : X \to 2^Y$. Multifunctions will be employed in various parts of this thesis as well as in many ways. While drawing mainly upon [51] below we shall sum up results that will be of the most use.

First we recall the notion of Hausdorff distance. Let X be a metric space endowed with a metric $d : X \times X \to \mathbb{R}_+$. For two arbitrary sets $A, B \subset X$ we define the Hausdorff distance:

$$D_H(A,B) := \max \left\{ \sup_{x \in A} d_B(x)\,,\, \sup_{y \in B} d_A(y) \right\} \tag{2.15}$$

where

$$d_B(x) := \inf_{y \in B} d(x,y), \qquad d_A(y) := \inf_{x \in A} d(x,y).$$

Let us recall that an *extended metric* is a function that satisfies all the properties of a metric except that it may admit infinite values.

Proposition 2.28. *Let $\mathcal{F}(X)$ be the set of all closed subsets of X. Then $\big(\mathcal{F}(X), D_H\big)$ is an extended metric space.*

Proof. See [51, Section II.1]. □

We say that a sequence of closed sets $\{C_n\} \subset \mathcal{F}(X)$ converges to $C \in \mathcal{F}(X)$ *in the sense of Kuratowski* when the two conditions are met:

$$\begin{cases} x_n \in C_n \quad \Rightarrow \quad \text{every cluster point of } \{x_n\} \text{ in } X \text{ belongs to C,} \\ \forall\, x \in C \text{ there exist } x_n \in C_n \text{ such that } x_n \to x \text{ in } X. \end{cases}$$

It is fairly straightforward to check that:

Proposition 2.29. *Assume that (X, d) is a compact metric space. Then $\big(\mathcal{F}(X), D_H\big)$ is a compact metric space whose topology induces the Kuratowski convergence.*

We may readily move on to summarize some topological properties of multifunctions $\Gamma : X \to 2^Y$. Naturally, by the graph of Γ we shall understand a subset of the Cartesian product $X \times Y$ as follows: $\mathcal{G}_\Gamma := \big\{(x, y) \,:\, x \in X,\ y \in \Gamma(x)\big\}$.

Definition 2.30 (**Continuity properties of multifunctions**)**.** Let Γ be a multifunction from a topological space X to subsets of a topological space Y.

(i) We say that Γ is lower semi-continuous at $x_0 \in X$ if for every open set $\mathscr{V}$ in Y that intersects with $\Gamma(x_0)$ there exists a neighbourhood $\mathscr{U}$ of x_0 such that

$$\Gamma(x) \cap \mathscr{V} \neq \varnothing \qquad \forall\, x \in \mathscr{U}.$$

(ii) We say that Γ is upper semi-continuous at $x_0 \in X$ if for every open set $\mathscr{V}$ in Y that contains $\Gamma(x_0)$ there exists a neighbourhood $\mathscr{U}$ of x_0 such that

$$\Gamma(x) \subset \mathscr{V} \qquad \forall\, x \in \mathscr{U}.$$

(iii) We say that Γ is continuous at $x_0 \in X$ if it is both lower and upper semi-continuous at x_0.

Moreover, we shortly say that $\Gamma : X \to 2^Y$ is lower semi-continuous (resp. upper semi-continuous or continuous) if it is lower semi-continuous (resp. upper semi-continuous or continuous) at every point in X.

It is always true that an upper semi-continuous multifunction Γ has a graph $\mathcal{G}_\Gamma$ that is closed. Once the space Y is compact we have the converse result as well:

Theorem 2.31 (**Multifunction's upper semi-continuity criterion**)**.** *Let Γ be a multifunction from a topological space X to subsets of a topological space Y. Assume that Y is a compact. Then Γ is upper semi-continuous if and only if its graph $\mathcal{G}_\Gamma$ is closed in the product topological space $X \times Y$.*

Proof. See [112]. □

Plenty of additional results may be established when Γ has values being convex subsets of a linear topological space Y. Below we give a criterion for lower semi-continuity of Γ once Y is finite dimensional:

Theorem 2.32 (**Multifunction's lower semi-continuity criterion**)**.** *Let Γ be a multifunction from a topological space X to convex subsets of a finite dimensional space Y.*

Then Γ is lower semi-continuous if and only if the mapping $X \times Y \ni (x, y) \mapsto \chi^*_{\Gamma(x)}(y)$ is lower semi-continuous, where $\chi^*_{\Gamma(x)}$ is the support function of $\Gamma(x)$.

Proof. See [38, Lemma A2]. □

It is natural that from $\Gamma : X \to 2^Y$ we wish to extract a so called *selection*, i.e. a univalued function $f : X \to Y$ satisfying $f(x) \in \Gamma(x)$ for each x. We reach the celebrated result by Michael on the sufficient condition of existence of selections that are continuous:

Theorem 2.33 (**Michael continuous selection theorem**). *Let Γ be a lower semi-continuous multifunction from a metric space X to non-empty closed convex subsets of a Banach space Y. Then there exists a continuous selection of Γ, i.e. a continuous function $f : X \to Y$ such that $f(x) \in \Gamma(x)$ for every $x \in X$.*

Proof. See [126, Theorem 3.2"]. □

Next we make a recollection of the notion of measurability for multifunctions. This issue is delicate as there are plenty of definitions depending both on the topological structure of the target space Y and on the properties of set values $\Gamma(x)$. Below we propose one of the definitions and then we give another one that (in the chosen setting) is equivalent. We agree that (X, Σ) will stand for a measurable space where Σ is a σ-algebra of measurable subsets of X.

Definition 2.34 (**Measurability of multifunctions**). Let Γ be a multifunction from a measurable space (X, Σ) to subsets of a separable metric space Y.

We say that Γ is measurable if for every open set $\mathscr{V}$ in Y it holds that

$$\Big\{x \in X \,:\, \Gamma(x) \cap \mathscr{V} \neq \varnothing\Big\} \;\in\; \Sigma.$$

Theorem 2.35 (**Measurability of multifunctions – alternative definition**). *For Γ as in Definition 2.34 assume that $\Gamma(x)$ is compact for every x. Then Γ is measurable if and only if the mapping*

$$X \ni x \;\mapsto\; \Gamma(x) \in \mathcal{F}(Y)$$

is measurable in the classical sense for $\mathcal{F}(Y)$ being equipped with D_H metric.

Proof. In [51] see Definition 1 in Secton III.1 and then the remark below Definition 10 in Section III.2. □

The notion of measurability of a multifunction is designed so that a measurable selection may be chosen:

Theorem 2.36 (**Measurable selection theorem**). *Let Γ be a measurable multifunction from a measurable space (X, Σ) to complete non-empty subsets of a separable metric space Y. Then there exists a Σ-measurable selection of Γ, i.e. a Σ-measurable function $f : X \to Y$ such that $f(x) \in \Gamma(x)$ for every $x \in X$.*

Proof. See [51, Theorem III.6]. □

One of the efficient ways to check whether a multifunction is measurable with respect to Borel σ-algebra $\Sigma = \mathscr{B}(X)$ is to show its upper semi-continuity:

Theorem 2.37 (**Multifunction's Borel measurability criterion**). *Let Γ be a multifunction from a topological space X to subsets of a metric space Y. Then, if Γ is upper semi-continuous (or lower semi-continunous) it is measurable as multifunction from a meaurable space $\big(X, \mathscr{B}(X)\big)$ to subsets of Y, where $\mathscr{B}(X)$ is the Borel σ-algebra.*

Proof. See [51, Corollary III.3]. □

Finally we give a criterion of μ-measurabiility (see Section 2.2 below) that appeals to the measurability of the graph:

Theorem 2.38 (**Multifunction's μ-measurability criterion**). *For a Radon measure μ on a locally compact separable metric space X let $\mathscr{B}_\mu(X)$ be the μ-completion of the Borel σ-algebra $\mathscr{B}(X)$. Let us consider a multifunction from X to non-empty closed subsets of a complete separable metric space Y.*

Then Γ is measurable with respect to the measurable space $\big(X, \mathscr{B}_\mu(X)\big)$ if and only if the graph $\mathcal{G}_\Gamma$ is measurable with respect to the product σ-algebra $\mathscr{B}_\mu(X) \otimes \mathscr{B}(Y)$.

Proof. See [51, Theorem III.30]. □

2.4. Maximal monotone maps on $\mathbb{R}^d$

From the paper [3] on maximal monotone maps we shall now gather results that will prove fundamental in Chapter 5. We say that a multifunction $\Gamma \equiv \mathbf{v} : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ is monotone whenever

$$\langle y' - x', y - x\rangle \geq 0 \qquad \forall\, (x,y) \in \mathbb{R}^d \times \mathbb{R}^d, \quad \forall\, x' \in \mathbf{v}(x),\ \forall\, y' \in \mathbf{v}(y).$$

The domain $\operatorname{dom}\mathbf{v}$ of the multifunction $\mathbf{v}$ will be the subset of $\mathbb{R}^d$ where $\mathbf{v}$ is non-empty valued, i.e. $\operatorname{dom}\mathbf{v} := \big\{x \in \mathbb{R}^d : \mathbf{v}(x) \neq \varnothing\big\}$. We will say that $\mathbf{v}$ *is of full domain* when $\operatorname{dom}\mathbf{v} = \mathbb{R}^d$.

For each $\mathbf{v}$ by $\mathcal{G}_\mathbf{v} \subset \mathbb{R}^d \times \mathbb{R}^d$ we shall understand its graph. A monotone multifunction $\mathbf{v}$ will be called *a maximal monotone map* on $\mathbb{R}^d$ provided that

$$\mathcal{G}_{\mathbf{v}'} \supset \mathcal{G}_\mathbf{v} \quad \text{for } \mathbf{v}' \text{ monotone} \qquad \Rightarrow \qquad \mathbf{v}' = \mathbf{v}.$$

Below we gather some basic properties of maximal monotone maps:

Proposition 2.39. *For every maximal monotone map $\mathbf{v}$ the following facts hold true:*

(i) for every $x \in \mathbb{R}^d$ the set $\mathbf{v}(x)$ is closed and convex;

(ii) the graph $\mathcal{G}_\mathbf{v}$ is closed, namely

$$x_n \to x,\ x'_n \to x' \quad \text{where } x'_n \in \mathbf{v}(x_n) \qquad \Rightarrow \qquad x' \in \mathbf{v}(x);$$

(iii) the domain of $\mathbf{v}$ contains the interior of its convex hull, i.e.

$$\operatorname{int}\Big(\operatorname{co}\big(\operatorname{dom}\mathbf{v}\big)\Big) \subset \operatorname{dom}\mathbf{v};$$

(iv) the multifunction $\mathbf{v}$ is upper semi-continuous in the interior of its domain $\operatorname{dom}\mathbf{v}$.

Proposition 2.40 (**Maximality criterion**). *Assume a monotone multifunction* $\mathbf{v}$ *of full domain, i.e.* $\operatorname{dom}\mathbf{v} = \mathbb{R}^d$*, that is upper semi-continuous and convex valued. Then* $\mathbf{v}$ *is a maximal monotone map.*

Proof. See [3, Corollary 1.4]. □

In the sequel of this section by $\mathbf{M}$ we will denote the set of maximal monotone maps on $\mathbb{R}^d$. Let us agree that by $\emptyset : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ we will understand an empty valued multifunction, i.e. $\emptyset(x) = \varnothing$ for all $x \in \mathbb{R}^d$. Note that $\emptyset$ is trivially a monotone multifunction although it is not maximal.

We say that a sequence of multifunctions $(\mathbf{v}_n)$ converges to $\mathbf{v}$ *in the sense of graph* when the sequence of graphs $\mathcal{G}_{\mathbf{v}_n}$ converges to $\mathcal{G}_{\mathbf{v}}$ in the sense of Kuratowski. It may be easily checked that the graph convergence amounts to verifying the two conditions:

$$\begin{cases} x_n \in \mathbb{R}^d,\ x'_n \in \mathbf{v}_n(x_n) \quad \Rightarrow \quad \text{every cluster point } (x, x') \text{ of } \{(x_n, x'_n)\} \text{ satisfies } x' \in \mathbf{v}(x), \\ \forall\, x \in \mathbb{R}^d,\ x' \in \mathbf{v}(x) \text{ there exist } x_n \in \mathbb{R}^d,\ x'_n \in \mathbf{v}_n(x_n) \text{ such that } (x_n, x'_n) \to (x, x'). \end{cases} \tag{2.16}$$

The following compactness result holds true for the set of maximal monotone maps:

Theorem 2.41. *On the set* $\mathbf{M} \cup \{\emptyset\}$ *there exists an extended metric* $\mathbf{d}$ *such that:*

(i) $\big(\mathbf{M} \cup \{\emptyset\}, \mathbf{d}\big)$ *is a compact metric space;*
(ii) the metric $\mathbf{d}$ *induces the graph convergence on* $\mathbf{M} \cup \{\emptyset\}$*.*

Moreover, assume a subset of $\mathbf{M}_0 \subset \mathbf{M}$ *that is closed in* $\big(\mathbf{M} \cup \{\emptyset\}, \mathbf{d}\big)$*. Then the topology on* $\mathbf{M}_0$ *is equivalently induced by the extended metric* $\mathbf{h} : \mathbf{M}_0 \times \mathbf{M}_0 \to [0, \infty]$ *defined as the Hausdorff distance between the graphs*

$$\mathbf{h}(\mathbf{v}_1, \mathbf{v}_2) := D_H\big(\mathcal{G}_{\mathbf{v_1}}, \mathcal{G}_{\mathbf{v_2}}\big).$$

Proof. See [3, Proposition 1.7]. □

As a consequence of Theorem 2.41, for any sequence of maximal monotone maps $\{\mathbf{v}_n\} \subset \mathbf{M}$ two scenarios are possible: (1) either there exists a subsequence $\{\mathbf{v}_{n_k}\}$ such that $\mathbf{v}_{n_k} \to \mathbf{v} \in \mathbf{M}$ in the sense of graph or (2) the sequence $\{\mathbf{v}_n\}$ diverges to an empty multifunction, i.e. for any compact set $K \subset \mathbb{R}^d$ it holds that

$$\liminf_{n\to\infty}\ \inf\Big\{|x'| \,:\, x \in K,\ x' \in \mathbf{v}_n(x)\Big\} \ =\ \infty.$$

Clearly for a set $\mathbf{M}_0$ satisfying the prerequisite of Theorem 2.41 only the possibility (1) remains as $\mathbf{M}_0$ is isolated from $\emptyset$.

For a maximal monotone map $\mathbf{v}$ we define a set of points x where the dimension of the convex set $\mathbf{v}(x)$ is at least $k \in \{1, \dots, d\}$:

$$\Sigma^k(\mathbf{v}) := \Big\{x \in \mathbb{R}^d \,:\, \dim \mathbf{v}(x) \geq k\Big\}$$

Moreover we say that $\mathbf{v}$ is *univalued at* x when $\mathbf{v}(x)$ is a singleton; we define the set

$$\mathrm{U}(\mathbf{v}) := \Big\{x \in \mathbb{R}^d \,:\, \mathbf{v}(x) = \{x'\},\ x' \in \mathbb{R}^d\Big\}.$$

Theorem 2.42. *For any maximal monotone map* $\mathbf{v}$ *the Hausdorff dimension of the set* $\Sigma^k(\mathbf{v})$ *is at most* $(d-k)$*. In particular the set* $\Sigma^1(\mathbf{v})$ *is of zero Lebesgue measure and as a consequence* $\mathbf{v}$ *is univalued* $\mathcal{L}^d$*-a.e. in the domain* $\operatorname{dom}\mathbf{v}$*, i.e.*

$$\mathcal{L}^d\Big(\big(\operatorname{dom}\mathbf{v}\big)\backslash\mathrm{U}(\mathbf{v})\Big)=0.$$

Proof. See [3, Theorem 2.2]. □

Theorem 2.43. *A maximal monotone map* $\mathbf{v}$ *is differentiable at* $\mathcal{L}^d$*-a.e. point* $x_0\in\mathrm{U}(\mathbf{v})$ *in the sense that there exists a matrix* $\nabla\mathbf{v}(x_0)\in\mathbb{R}^{d\times d}$ *such that*

$$\lim_{\substack{x\to x_0\\ x'\in\mathbf{v}(x)}}\frac{x'-\mathbf{v}(x_0)-\nabla\mathbf{v}(x_0)(x-x_0)}{|x-x_0|}=0. \tag{2.17}$$

Proof. See [3, Theorem 3.2]. □

Corollary 2.44. *A maximal monotone map* $\mathbf{v}$ *is differentiable in the sense of* (2.17) *for* $\mathcal{L}^d$*-a.e.* $x_0\in\operatorname{dom}\mathbf{v}$*.*

Thanks to the results above and to the fact that each $\mathbf{v}\in\mathbf{M}$ is an upper semi-continuous multifunction in the interior of $\operatorname{dom}\mathbf{v}$, we may assign to it a Borel measurable function v defined therein. The next result discusses the distributional derivative Dv of such function:

Theorem 2.45. *Let* $\mathscr{U}$ *be an open convex set in* $\mathbb{R}^d$*.*

(i) For any maximal monotone map $\mathbf{v}$ *such that* $\mathscr{U}\subset\operatorname{dom}\mathbf{v}$ *there exists* $v\in BV_{\mathrm{loc}}\cap L^\infty_{\mathrm{loc}}(\mathscr{U};\mathbb{R}^d)$ *such that* $\mathbf{v}(x)=\big\{v(x)\big\}$ *for a.e.* $x\in\mathscr{U}$ *while*

$$\varepsilon(v):=\frac{1}{2}\Big(\mathrm{D}v+(\mathrm{D}v)^\top\Big) \tag{2.18}$$

is a positive matrix valued measure, i.e. $\varepsilon(v)\in\mathcal{M}(\mathscr{U};\mathcal{S}^{d\times d}_+)$*. Moreover, for any open set* $\mathscr{V}\Subset\mathscr{U}$

$$\int_{\mathscr{V}}|\mathrm{D}v|\le\Big(C_d\big(\operatorname{diam}(\mathscr{V})\big)\Big)^{d-1}\operatorname{osc}(\mathbf{v},\mathscr{V})$$

where $\operatorname{diam}(\mathscr{V})$ *is the diameter of* $\mathscr{V}$ *and* $\operatorname{osc}(\mathbf{v},\mathscr{V}):=\sup\big\{|y'-x'|:(x,y)\in\mathscr{V}\times\mathscr{V},\ x'\in\mathbf{v}(x),\ y'\in\mathbf{v}(y)\big\}$*.*

(ii) Conversely, assume that $v\in L^\infty_{\mathrm{loc}}(\mathscr{U};\mathbb{R}^d)$ *and that* $\varepsilon(v)$ *defined in* (2.18) *is a positive matrix valued distribution. Then there exists a maximal monotone map* $\mathbf{v}$ *such that* $\mathbf{v}(x)=\big\{v(x)\big\}$ *for a.e.* $x\in\mathscr{U}$*. Therefore* $v\in BV_{\mathrm{loc}}(\mathscr{U};\mathbb{R}^d)$ *by the assertion (i).*

Proof. See [3, Proposition 5.1, Remark 5.2, Theorem 5.3]. □

2.5. The μ-tangential calculus

It is a natural direction to attempt extending the study of Poisson's equation $-\triangle u=f$ posed within domain $\Omega\subset\mathbb{R}^d$ to its counterpart on a positive measure $\mu\in\mathcal{M}^b_+(\mathbb{R}^d)$, especially if the latter describes a junction of lower dimensional geometric objects, such as manifolds of different dimensions. Formally such equation may be written down as $-\operatorname{div}(\nabla u\,\mu)=f$ where f is for instance an element in $L^2_\mu(\mathbb{R}^d)$. It soon becomes clear that

the solution cannot be an element of standard Sobolev space $W^{1,2}_{\mathrm{loc}}(\mathbb{R}^d)$ since the gradient ∇u must be defined μ-a.e. In the pioneering work [40] a new functional space $W^{1,p}_\mu$ was put forward for which the conventional weak gradient is replaced by the *μ-tangential gradient* $\nabla_\mu u$. For smooth u we have $\nabla_\mu u = P_\mu \nabla u$ where point-wisely $P_\mu(x)$ is the orthogonal projection onto the space $T_\mu(x) \subset \mathbb{R}^d$ tangent to measure at x, whose definition is intrinsically linked to the definition of $W^{1,p}_\mu$ itself. Since the original paper [40] the theory of μ-tangential calculus evolved greatly, including generalization to differential operators other than ∇, see e.g. [86, 35, 36, 37]. Recently this theory was revisited in [150] and also by the present author in [27].

Below we present an overview of results on the μ-tangential calculus that will be of the most use in this thesis. We shall combine results from the papers quoted above. We will give proofs for most of them in order to be as self-contained as possible and also to slightly adjust some of them to our needs.

2.5.1. The abstract μ-tangential operator A_μ

Let us fix two finite dimensional spaces U, V, an open set $\mathscr{U} \subset \mathbb{R}^d$ and a positive finite Radon measure $\mu \in \mathcal{M}^b_+(\mathscr{U})$. On top of that we take a linear continuous operator

$$A : \mathcal{D}(\mathscr{U}; U) \to C_0(\mathscr{U}; V)$$

where continuity is intended with respect to the standard topology on the space of test functions $\mathcal{D}(\mathscr{U}; U)$. Further, for the linear space $\mathcal{U} = \mathcal{D}(\mathscr{U}; U)$ and a norm $|||\cdot|||$ of our choice, by Banach space $\overline{\mathcal{U}}$ we will understand the completion of $\mathcal{U}$ in this very norm. Finally, we set an exponent $q \in [1, \infty]$ whereas q' shall stand for its Hölder conjugate.

Example 2.46. The theory herein recalled is designed for A being a differential operator. The easiest case would consist in taking: $q < \infty$, $U = \mathbb{R}$, $V = \mathbb{R}^d$ and $A = \nabla$ with $|||\cdot||| = \|\cdot\|_{L^q_\mu}$ yielding $\overline{\mathcal{U}} = L^q_\mu(\mathscr{U}; \mathbb{R})$. △

For a graph of the operator A restricted to domain $\mathcal{U}$:

$$\mathcal{G} := \Big\{ (u, Au) \,:\, u \in \mathcal{U} \Big\}$$

we take its closure in the space $\overline{\mathcal{U}} \times L^q_\mu(\mathscr{U}; V)$, more specifically

$$\overline{\mathcal{G}}^{\,q} := \Big\{ (u, \epsilon) \in \overline{\mathcal{U}} \times L^q_\mu(\mathscr{U}; V) \;:\; \exists \{u_n\} \subset \mathcal{U}, \quad |||u_n - u||| \to 0,$$
$$Au_n \to \epsilon \ \text{ in } L^q_\mu(\mathscr{U}; V) \ \Big(Au_n \overset{*}{\rightharpoonup} \epsilon \ \text{ in } L^\infty_\mu(\mathscr{U}; V) \ \text{ for } \ q = \infty \Big) \Big\}.$$

Remark 2.47. We point out that for $q < \infty$ the strong convergence in definition of $\overline{\mathcal{G}}$ may be equivalently replaced by weak convergence in L^q: it comes from the fact that $\mathcal{G}$ is a linear space while in locally convex topological spaces norm and weak closures of convex sets are identical, cf. [147]. Considering boundedness of μ this brings us to the following conclusion:

$$\overline{\mathcal{G}}^{\,q_2} \subset \overline{\mathcal{G}}^{\,q_1} \qquad \text{for } 1 \le q_1 \le q_2 \le \infty.$$

In what follows all the introduced notions will be sensitive to the exponent q – we shall suppress this dependence in notation. In general the linear space $\overline{\mathcal{G}} = \overline{\mathcal{G}}^{\,q}$ is not a graph, for the space

$$\mathcal{N} := \left\{ \epsilon \in L^q_\mu(\mathscr{U}; V) \ : \ (0, \epsilon) \in \overline{\mathcal{G}} \right\} \tag{2.19}$$

is non-trivial. The underlying idea (in the framework of Example 2.46) is that the vector valued functions in $\mathcal{N}$ point-wisely μ-a.e. span the space orthogonal to measure μ. By computing the orthogonal complement of $\mathcal{N}$ we recover the space of functions that are "tangent to measure" in the spirit of the pioneering paper [40]:

Proposition 2.48. *The set $\mathcal{N}$ is a closed subspace of $L^q_\mu(\mathscr{U}; V)$ while*

$$\mathcal{N}^\perp = \mathrm{cl} \left\{ \sigma \in L^{q'}_\mu(\mathscr{U}; V) \ : \ \exists\, C > 0 \text{ such that } \left| \int \langle Au, \sigma \rangle \, d\mu \right| \le C |||u||| \quad \forall\, u \in \mathcal{U} \right\}$$

where the closure is intended in the strong sense in $L^{q'}_\mu(\mathscr{U}, V)$ for $q \in (1, \infty]$ and in the weak- sense in $L^\infty_\mu(\mathscr{U}; V)$ for $q = 1$.*

Proof. It is clear that $\mathcal{N}$ is a subspace that is sequentially closed in $L^q_\mu(\mathscr{U}; V)$: in norm topology when $q < \infty$ and in weak-* topology when $q = \infty$. Topological closedness is trivial when $q < \infty$, meanwhile it follows from Banach-Dieudonne theorem (see [83, Corollary 3.93]) when $q = \infty$.

With $(\overline{\mathcal{U}})^*$ being the dual to Banach space $\overline{\mathcal{U}}$ we can characterize the orthogonal complement

$$\mathcal{G}^\perp = \left\{ (u^*, \sigma) \in (\overline{\mathcal{U}})^* \times L^{q'}_\mu(\mathscr{U}, V) \ : \ \langle u, u^* \rangle + \int \langle Au, \sigma \rangle \, d\mu = 0 \quad \forall\, u \in \mathcal{U} \right\}.$$

Let $\mathcal{T} \subset L^{q'}_\mu(\mathscr{U}, V)$ be the projection of $\mathcal{G}^\perp$ onto the second coordinate. Since $\overline{\mathcal{G}} = (\mathcal{G}^\perp)^\perp$ we can check that $\mathcal{T}^\perp = \mathcal{N}$ or $\overline{\mathcal{T}} = \mathcal{N}^\perp$. Thus σ is an element of $\mathcal{T}$ if and only if there exists $u^* \in (\overline{\mathcal{U}})^*$ such that $\int \langle Au, \sigma \rangle \, d\mu = -\langle u, u^* \rangle$, which furnishes the assertion. □

The space $\mathcal{N}$ encodes the global information on the orthogonality to measure μ whilst for localizing we require an additional *stability assumption*:

$$\epsilon \in \mathcal{N}, \quad \psi \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}) \qquad \Rightarrow \qquad \psi\epsilon \in \mathcal{N}. \tag{$\mathrm{H}_{\mathrm{stab}}$}$$

Example 2.49. Verification of ($\mathrm{H}_{\mathrm{stab}}$) is straightforward in the setting of Example 2.46: for $\epsilon \in \mathcal{N}$ we have by definition a sequence $\{u_n\} \subset \mathcal{U}$ such that $\|u_n\|_{L^q_\mu} \to 0$ and $Au_n = \nabla u_n \to \epsilon$ in L^q_μ; then for $\tilde{u}_n = \psi\, u_n \in \mathcal{U}$ again $\|\tilde{u}_n\|_{L^q_\mu} \to 0$ while $A\tilde{u}_n = u_n \nabla\psi + \psi \nabla u_n$ converges to $\psi\epsilon$ in L^q_μ rendering it an element of $\mathcal{N}$. △

Assumption ($\mathrm{H}_{\mathrm{stab}}$) lies closely to notion of *decomposability* (cf. [137]) paving a way to the desirable local result:

Proposition 2.50. *Under stability condition* ($\mathrm{H}_{\mathrm{stab}}$) *there exists a μ-measurable multi-function N_μ from $\mathscr{U}$ to linear subspaces of V such that for any $\epsilon \in L^q_\mu(\mathscr{U}; V)$ and any $\sigma \in L^{q'}_\mu(\mathscr{U}; V)$ there hold the equivalences*

$$\begin{aligned} \epsilon \in \mathcal{N} \qquad &\Leftrightarrow \qquad \epsilon(x) \in N_\mu(x) \quad \text{for } \mu\text{-a.e. } x, \\ \sigma \in \mathcal{N}^\perp \qquad &\Leftrightarrow \qquad \sigma(x) \in N^\perp_\mu(x) \quad \text{for } \mu\text{-a.e. } x. \end{aligned}$$

Proof. See [37, Proposition 4]. □

Condition (H_{stab}) shall be assumed throughout the current subsection. Readily, for μ-a.e. x, $N_\mu(x)$ is a normal space to μ at a point. Consequently, the subspace $N_\mu^\perp(x)$ deserves to be treated as tangential to μ at x. We are in a position to define the μ-tangential operator:

Definition 2.51 (**The μ-tangential operator A_μ**)**.** For μ-a.e. x let

$$\Pi_\mu(x) : V \to M_\mu(x) := N_\mu^\perp(x)$$

be the orthogonal projection onto the *tangential subspace* $M_\mu(x) \subset V$. We introduce the linear operator $A_\mu : \mathrm{dom}(A_\mu) \to L_\mu^q(\mathscr{U}; V)$ with the domain

$$\mathrm{dom}(A_\mu) = \mathrm{dom}(A_\mu)(\mathscr{U}, q, |||\cdot|||) := \left\{ u \in \overline{\mathcal{U}} \ : \ \exists\, \epsilon \in L_\mu^q(\mathscr{U}; V),\ (u, \epsilon) \in \overline{\mathcal{G}} \right\}$$

that is defined via

$$A_\mu u = \Pi_\mu \epsilon$$

where ϵ is any element of $L_\mu^q(\mathscr{U}; V)$ such that $(u, \epsilon) \in \overline{\mathcal{G}}$.

Remark 2.52. As an orthogonal projection the operator $\Pi_\mu(x)$ is symmetric, namely $\langle \Pi_\mu(x)\xi, \sigma\rangle = \langle \xi, \Pi_\mu(x)\sigma\rangle$ for any $\xi, \sigma \in V$.

Remark 2.53. For two open sets satisfying $\mathscr{U}_1 \subset \mathscr{U}_2$ there clearly holds $\mathrm{dom}(A_\mu^1)(\mathscr{U}_1, q, |||\cdot|||) \subset \mathrm{dom}(A_\mu^2)(\mathscr{U}_2, q, |||\cdot|||)$. Moreover for any $u \in \mathrm{dom}(A_\mu^1)(\mathscr{U}_1, q, |||\cdot|||)$ we have $A_\mu^1 u = A_\mu^2 u$ μ-a.e. in $\mathscr{U}_1$.

As already noted, $\overline{\mathcal{G}}$ is in general not a graph of a univalued linear map, i.e. for given $u \in \mathrm{dom}(A_\mu)$ there are multiple $\epsilon \in L_\mu^q(\mathscr{U}; V)$ satisfying $(u, \epsilon) \in \overline{\mathcal{G}}$. However, thanks to the construction above, ϵ is unique up to an element in $\mathcal{N}$ hence the point-wisely computed $\Pi_\mu \epsilon$ is unique by virtue of Proposition 2.50 and A_μ is ultimately well defined.

The operator A_μ may be equivalently defined through closure of an unbounded operator – the graph of A_μ is the $\overline{\mathcal{U}} \times L_\mu^q(\mathscr{U}; V)$-closure of an operator defined only on domain $\mathcal{U}$:

Proposition 2.54. *The operator A_μ is the closure of the unbounded operator $\tilde{A}_\mu : \mathcal{U} \subset \overline{\mathcal{U}} \to L_\mu^q(\mathscr{U}; V)$ given by*

$$\tilde{A}_\mu u := \Pi_\mu(Au) \qquad \forall\, u \in \mathcal{U}.$$

As a consequence for $q < \infty$ the domain $\mathrm{dom}(A_\mu)$ is a Banach space being the completion of $\mathcal{U}$ with respect to the graph norm

$$\|u\|_{\mathrm{dom}(A_\mu)} := |||u||| + \|A_\mu u\|_{L_\mu^q}.$$

Proof. We must show that for any sequence $\{u_n\} \subset \mathcal{U}$ such that there exists a pair $(u, \epsilon) \in \overline{\mathcal{U}} \times L_\mu^q(\mathscr{U}; V)$ for which $u_n \to u$ in $\overline{\mathcal{U}}$ and $\tilde{A}_\mu u_n \to \epsilon$ in $L_\mu^q(\mathscr{U}; V)$ ($\tilde{A}_\mu u_n \overset{*}{\rightharpoonup} \epsilon$ if $q = \infty$) we have $u \in \mathrm{dom}(A_\mu)$ and $A_\mu u = \epsilon$, or equivalently that $(u, \epsilon) \in \overline{\mathcal{G}}$. For each n from Proposition 2.50 we see that $\Pi_\mu^\perp(Au_n) := Au_n - \Pi_\mu(Au_n) \in \mathcal{N}_\mu$ thus $\big(0, \Pi_\mu^\perp(Au_n)\big) \in \overline{\mathcal{G}}$; clearly $(u_n, Au_n) \in \overline{\mathcal{G}}$ as well. Then, since $\overline{\mathcal{G}}$ is a linear space

$$\big(u_n, \tilde{A}_\mu u_n\big) = \big(u_n, \Pi_\mu(Au_n)\big) = \big(u_n, Au_n\big) - \big(0, \Pi_\mu^\perp(Au_n)\big) \in \overline{\mathcal{G}}.$$

The space $\overline{\mathcal{G}}$ is closed by definition and the claim follows. □

As a corollary we can give a useful criterion for checking whether a function u lies in the domain of the operator A_μ:

Corollary 2.55. *Assume that $q \in (1,\infty]$ and consider a function $u \in \overline{\mathcal{U}}$. If there exists a sequence $u_n \in \mathcal{U}$ such that:*

(i) $|||u_n - u||| \to 0$,
(ii) $\sup_n \|A_\mu u_n\|_{L^q_\mu(\mathscr{U};V)} < \infty$,

then $u \in \mathrm{dom}(A_\mu)$.

Proof. Thanks to reflexiveness of L^q_μ there exists ϵ such that (up to extraction of a subsequence) $A_\mu u_n \to \epsilon$ in L^q_μ. The assertion follows by the closedness of the graph of the operator A_μ. □

We find that the tangential space $M_\mu(x)$ is local in the sense that it depends only on restriction of measure μ to a ball centred at x. By exploiting the stability property ($\mathrm{H}_{\mathrm{stab}}$) it is straightforward to deduce the following:

Proposition 2.56 (**Locality of the tangential space**)**.** *Let $\mu,\nu \in \mathcal{M}^b_+(\mathscr{U})$ be two bounded Radon measures and let $\mathscr{V}$ be any open subset of $\mathscr{U}$. Then the following implication holds:*

$$\mu \llcorner \mathscr{V} = \nu \llcorner \mathscr{V} \qquad \Rightarrow \qquad M_\mu = M_\nu \quad \mu\text{-a.e. in } \mathscr{V}.$$

We reach the main result of this presentation: the μ-tangential integration by parts formula. We take a distribution $\mathscr{F} \in \mathcal{D}'(\mathscr{U};U)$ for which there exists a constant $C \geq 0$ satisfying

$$|\mathscr{F}(u)| \leq C|||u||| \qquad \forall\, u \in \mathcal{U} = \mathcal{D}(\mathscr{U};U). \tag{2.20}$$

An immediate implication of this bound is existence of an extension of $\mathscr{F}$ to a continuous linear functional $\mathscr{F} : \overline{\mathcal{U}} \to \mathbb{R}$ (we do not relabel). Next, let A^* stand for the adjoint of A, with A treated as an operator from $\mathcal{D}(\mathscr{U};U)$ to $C_0(\mathbb{R}^d;V) \supset C_0(\mathscr{U};V)$, i.e. we consider a linear continuous operator

$$A^* : \mathcal{M}^b(\mathbb{R}^d;V) \to \mathcal{D}'(\mathscr{U};U).$$

Theorem 2.57 (**Integration by parts formula for operator A_μ**)**.** *Let there be given a bounded Radon measure $\bar\mu \in \mathcal{M}^b_+(\mathbb{R}^d)$, a function $\sigma \in L^{q'}_{\bar\mu}(\mathbb{R}^d;V)$ and a distribution $\mathscr{F} \in \mathcal{D}'(\mathscr{U};U)$ satisfying* (2.20)*. Then, the distributional equation*

$$A^*(\sigma\bar\mu) = \mathscr{F} \qquad \text{in} \quad \mathcal{D}'(\mathscr{U};U)$$

implies that for the Radon measure $\mu = \bar\mu \llcorner \mathscr{U} \in \mathcal{M}^b_+(\mathscr{U})$ there hold:

(i) $\sigma(x) \in M_\mu(x)$ *for μ-a.e. $x \in \mathscr{U}$;*
(ii) the integration by parts formula

$$\int_{\mathscr{U}} \langle A_\mu u, \sigma\rangle \, d\mu = \mathscr{F}(u) \qquad \forall\, u \in \mathrm{dom}(A_\mu).$$

Proof. For any $\varphi \in \mathcal{D}(\mathscr{U}; U)$ we have $A\varphi \in C_0(\mathscr{U}; V)$ and thus

$$\int_{\mathscr{U}} \langle A\varphi, \sigma \rangle \, d\mu = \int_{\mathbb{R}^d} \langle A\varphi, \sigma \rangle \, d\bar{\mu} = \langle A\varphi, \sigma\bar{\mu} \rangle_{(C_0, \mathcal{M}^b)} = \langle \varphi, A^*(\sigma\bar{\mu}) \rangle_{(\mathcal{D}, \mathcal{D}')} = \mathscr{F}(\varphi) \leq C|||\varphi||| \tag{2.21}$$

therefore $\sigma|_{\mathscr{U}} \in \mathcal{N}^\perp$ owing to Proposition 2.48 and claim (i) follows by virtue of characterization in Proposition 2.50.

Let us now take a function $u \in \mathrm{dom}(A_\mu)$, then by definition we have at our disposal a sequence $u_n \in \mathcal{U} = \mathcal{D}(\mathscr{U}; U)$ and $\epsilon \in L^q_\mu(\mathscr{U}; V)$ such that $u_n \to u$ in $\overline{\mathcal{U}}$ and $Au_n \to \epsilon$ in $L^q_\mu(\mathscr{U}; V)$ ($Au_n \overset{*}{\rightharpoonup} \varepsilon$ if $q = \infty$) where $\Pi_\mu \epsilon = A_\mu u$. From (2.21) we thus obtain

$$\mathscr{F}(u) = \lim_{n\to\infty} \mathscr{F}(u_n) = \lim_{n\to\infty} \int_{\mathscr{U}} \langle Au_n, \sigma \rangle \, d\mu = \int_{\mathscr{U}} \langle \epsilon, \sigma \rangle \, d\mu = \int_{\mathscr{U}} \langle A_\mu u, \sigma \rangle \, d\mu$$

where we used the fact that $\langle \epsilon, \sigma \rangle = \langle \epsilon, \Pi_\mu \sigma \rangle = \langle \Pi_\mu \epsilon, \sigma \rangle$, which is due to (i) and Remark 2.52. □

2.5.2. The μ-tangential gradient, Sobolev space with respect to measure and the space tangent to measure

We start specifying the abstract theory of μ-tangential operator with a setting that was originally treated in the pioneering work [40]. For a bounded domain Ω that guarantees $\mathrm{spt}\,\mu \subset \overline{\Omega}$ we consider

$$\begin{cases} U = \mathbb{R}, \quad \mathscr{U} = \mathbb{R}^d, \quad \mathcal{U} = \mathcal{D}(\mathbb{R}^d), \\ |||\cdot||| = \begin{cases} \|\cdot\|_{L^q_\mu} & \text{if} \quad q \in [1, \infty), \\ \|\cdot\|_{\infty,\Omega} & \text{if} \quad q = \infty, \end{cases} \\ V = \mathbb{R}^d, \quad A = \nabla. \end{cases} \tag{2.22}$$

For $q < \infty$ this framework is designed to treat elliptic partial differential problems set on a given measure μ, e.g. the stationary heat conduction problem for $q = 2$, cf. [150]. This dissertation will be mainly concerned with the variant of $q = \infty$ that is specific to optimal design, i.e. the case when μ is unknown. In such scenario a feasible domain Ω must be fixed, which explains its presence in (2.22) only when $q = \infty$.

For $A = \nabla$ the stability condition ($\mathrm{H}_{\mathrm{stab}}$) has been already checked in Section 2.5.1, see Example 2.49 (for $q = \infty$ the proof is analogous). The μ-tangential calculus in the case of $A = \nabla$ is very much connected to the geometry of measure μ, which justifies some of the notation below:

$$A_\mu =: \nabla_\mu, \qquad \mathrm{dom}(A_\mu) =: W^{1,q}_\mu, \qquad \Pi_\mu =: P_\mu, \qquad M_\mu =: T_\mu$$

First, the domain of the operator ∇_μ deserves the name of *Sobolev space with respect to measure*: $W^{1,q}_\mu$, that is the space where the *μ-tangential gradient* can be defined and belongs to L^q_μ. Secondly, the space $T_\mu \subset \mathbb{R}^d$ ought to point-wise represent the intuitively understood space tangent to measure and P_μ is the respective orthogonal projection. To make this statement mathematically precise we consider a fairly wide class of geometric measures, see [40]:

Definition 2.58 (**Multi-junction measure**)**.** We shall say that measure μ is a *multi-junction measure* if it is of the form $\mu = \sum_{i=1}^{N} \mu_i$ where for each i: $\mu_i = m_i \mathcal{H}^{k_i} \llcorner S_i$ with m_i being a positive constant, k_i being an integer in $\{1, \ldots, d\}$ and S_i being a k_i-dimensional manifold of C^2 class. In addition we assume that $\mu_i(S_j) = 0$ whenever $i \neq j$.

Proposition 2.59. *For a multi-junction measure μ the tangent space enjoys the characterization:*

$$T_\mu = T_{\mu_i} = T_{S_i} \qquad \mu_i\text{-a.e.}$$

where T_{S_i} denotes the space tangent to manifold S_i in the classical sense.

Proof. See [36, Lemma 2.2] when $q < \infty$ and [86, Theorem 3.1] when $q = \infty$. □

Example 2.60. Let us take $\mu = \mathcal{L}^d \llcorner \Omega$ for a bounded domain Ω. Then, owing to the result above we have $T_\mu = \mathbb{R}^d$ a.e. in Ω. Moreover, for $q \in [1, \infty)$, the domain $W^{1,q}_\mu$ is identical to the classical Sobolev space $W^{1,q}(\Omega)$. △

Remark 2.61. From definition of $\overline{\mathcal{G}}^q$ it is easy to see that the above result holds true if m_i are positive weight functions satisfying $m_i \in L^\infty_{\mu_i}$ and $1/m_i \in L^\infty_{\mu_i}$. For weight functions m_i that are not separated from zero analysis of T_μ and the Sobolev space $W^{1,q}_\mu$ becomes delicate, even on the real line. The contribution of the author of this thesis to this topic involves a study of higher order spaces $W^{k,q}_\mu$ on a 1D interval where $\mu = m\,\mathcal{L}^1 \llcorner I$ for m that may degenerate to zero, cf. [27] – the main question concerns continuity of $u \in W^{k,q}_\mu$ and its derivatives at singular points of m.

Remark 2.62. In the framework (2.22) we may also consider any open subset $\mathscr{U} \subset \mathbb{R}^d$ and thus $\mathcal{U} = \mathcal{D}(\mathscr{U})$. Then the domain of the operator ∇_μ accounts "for some μ-intrinsic trace condition on $\partial\mathcal{U}$". For instance, if $q < \infty$ and $\mu = \mathcal{L}^d \llcorner \mathscr{U}$ we have $\mathrm{dom}(\nabla_\mu) = W^{1,q}_0(\mathscr{U})$ in the classical sense. Therefore, for arbitrary $\mathscr{U}$ the following symbol is in order:

$$\mathrm{dom}(\nabla_\mu) =: W^{1,q}_{\mu,0}(\mathscr{U}).$$

Remark 2.63. The framework examined in this section may be easily generalized to the case of $V = \mathbb{R}^m$, i.e. when functions u are $\mathbb{R}^m$-valued and $A = (\nabla)^m$, namely for $u = (u_1, \ldots, u_m)$ the operator Au is point-wisely a matrix whose columns are $\nabla u_1, \ldots, \nabla u_m$. In this setting the components of u are decoupled therefore

$$\mathrm{dom}(A_\mu) = \big(\mathrm{dom}(\nabla_\mu)\big)^m = \big(W^{1,q}_{\mu,0}(\mathscr{U})\big)^m, \qquad M_\mu = (T_\mu)^m, \qquad A_\mu = (\nabla_\mu)^m$$

where by $(T_\mu)^m$ we agree to understand a subspace of $d \times m$ matrices whose columns are elements of T_μ. In practice for functions $u \in \big(W^{1,q}_{\mu,0}(\mathscr{U})\big)^m$ we shall write $A_\mu u = \nabla_\mu u$ instead of $A_\mu u = (\nabla_\mu)^m u$.

The following μ-tangential chain rule will be of great importance in Chapter 6:

Proposition 2.64 (**The μ-tangential chain rule**)**.** *For $k, m \in \mathbb{N}_+$ let there be given a function $G \in C^1(\mathbb{R}^m; \mathbb{R}^k)$ such that the images of both G and $\nabla G \in C(\mathbb{R}^m; \mathbb{R}^{m\times k})$ are*

bounded. Then for every $u \in \big(W^{1,q}_{\mu,0}(\mathbb{R}^d)\big)^m$ *the composition* $v = G \circ u$ *is an element of* $\big(W^{1,q}_{\mu,0}(\mathbb{R}^d)\big)^k$ *and moreover*

$$\nabla_\mu v(x) = \Big(\nabla_\mu u(x)\Big)\Big(\nabla G\big(u(x)\big)\Big) \qquad \text{for } \mu\text{-a.e. } x \tag{2.23}$$

where the product is intended in the sense of classical multiplication of matrices.

Proof. Let us assume that $q < \infty$ (for $q = \infty$ the proof runs analogously). By the definition of $\big(W^{1,q}_{\mu,0}(\mathbb{R}^d)\big)^m$ we may point to a sequence $\{u_n\} \subset \mathcal{D}(\mathbb{R}^d;\mathbb{R}^m)$ such that $u_n \to u$ in $L^q_\mu(\mathbb{R}^d;\mathbb{R}^m)$ and $\nabla u_n \to \epsilon$ in $L^q_\mu(\mathbb{R}^d;\mathbb{R}^{d\times m})$ for some ϵ. For each n let us define $v_n = G \circ u_n$. Clearly $v_n \to v$ in $L^q_\mu(\mathbb{R}^d;\mathbb{R}^m)$ while by the classical chain rule we have for every $x \in \mathbb{R}^d$

$$\nabla v_n(x) = \Big(\nabla u_n(x)\Big)\Big(\nabla G\big(u_n(x)\big)\Big). \tag{2.24}$$

Owing to the prerequisite of the proposition $\nabla G\big(u_n(x)\big)$ is bounded in $\mathbb{R}^{m\times k}$ uniformly with respect to x and n. Hence (possibly upon extracting a subsequence) we may use the Lebesgue dominated convergence theorem to infer that $(\nabla u_n)\,(\nabla G \circ u_n)$ converges in $L^q_\mu(\mathbb{R}^d;\mathbb{R}^{d\times k})$ to $\zeta = \epsilon\,(\nabla G \circ u)$. As a result $\nabla v_n \to \zeta$ in $L^q_\mu(\mathbb{R}^d;\mathbb{R}^{d\times k})$ hence $v \in \big(W^{1,q}_{\mu,0}(\mathbb{R}^d)\big)^k$. We obtain (2.23) through left-sided multiplication of the equation $\zeta = \epsilon\,(\nabla G \circ u)$ by P_μ. □

The simplest class of functions that is contained in $W^{1,q}_{\mu,0}(\mathscr{U})$ for any q is the one of Lipschitz functions:

Proposition 2.65. *For an open set* $\mathscr{U} \subset \mathbb{R}^d$ *and measure* $\mu \in \mathcal{M}^b_+(\mathscr{U})$ *(of bounded support) let* $\Omega \subset \mathbb{R}^d$ *be any bounded domain such that* $\operatorname{spt}\mu \subset \overline{\Omega}$*. Then, for any* $q \in [1,\infty]$ *there holds an inclusion*

$$\operatorname{Lip}_0(\mathscr{U};\mathbb{R}^m) \quad \subset \quad \big(W^{1,q}_{\mu,0}(\mathscr{U})\big)^m$$

(note that for $q = \infty$ *the right hand side depends on* Ω*).*

Proof. In view of Remarks 2.47, 2.63 it is enough to show the claim for $q = \infty$ and $m = 1$. We take a function $u \in \operatorname{Lip}_0(\mathscr{U};\mathbb{R})$. Let us first assume that $\operatorname{spt} u \subset \mathscr{U}$. For $\varepsilon > 0$ let $u_\varepsilon(x) := (u * \eta_\varepsilon)(x)$ for $x \in \Omega$, where η_ε is the standard, radial-symmetric mollifying kernel satisfying $\operatorname{spt}\eta_\varepsilon = \overline{B}(0,\varepsilon)$. Then, for sufficiently small ε, we have $u_\varepsilon \in \mathcal{D}(\mathscr{U})$. Owing to [82, Section 4.2, Theorem 1 and Theorem 5] and to the averaging properties of molliffying we have: $u \in W^{1,\infty}(\mathbb{R}^d)$, $\nabla u_\varepsilon = \nabla u * \eta_\varepsilon$, $\|\nabla u_\varepsilon\|_{L^\infty(\overline{\Omega})} \le \|\nabla u\|_{L^\infty(\overline{\Omega})} < \infty$ and $u_\varepsilon \to u$ uniformly on $\overline{\Omega}$ when $\varepsilon \to 0$ ($\overline{\Omega}$ is a compact subset of $\mathbb{R}^d$). Accordingly, since $\operatorname{spt}\mu \subset \overline{\Omega}$, there exists a sequence $\{\varepsilon_n\}$ such that $\nabla u_{\varepsilon_n} \overset{*}{\rightharpoonup} \epsilon$ in $L^\infty_\mu(\mathscr{U};\mathbb{R}^d)$. As a result $u \in \operatorname{dom}(\nabla_\mu)$.

In the case when the support of u is not contained in $\mathscr{U}$, for $\delta > 0$ we define a function

$$u_\delta(x) := \begin{cases} u(x) - \delta & \text{if } u(x) \ge \delta, \\ u(x) + \delta & \text{if } u(x) \le -\delta, \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that u_δ is Lipschitz continuous while $\|\nabla u_\delta\|_{L^\infty(\overline{\Omega})} \le \|\nabla u\|_{L^\infty(\overline{\Omega})}$ and $u_\delta \to u$ uniformly on $\mathbb{R}^d$ when $\delta \to 0$. Moreover, since $u \in C_0(\mathscr{U};\mathbb{R})$ we deduce that

spt $u_\delta \subset \mathscr{U}$. Our claim follows by considering $u_{\delta,\varepsilon} = u_\delta * \eta_\varepsilon$ and by using the diagonalization argument. □

2.5.3. The μ-tangential strain operator

From Remark 2.63 we know that considering $V = \mathbb{R}^m$, i.e. $\mathbb{R}^m$-valued functions u, is not a real novelty if we stick to the operator ∇. The significantly new results emerge if for $m = d$ as the operator A we take the symmetric part of the gradient, i.e. if we consider the framework of elasticity:

$$\begin{cases} U = \mathbb{R}^d, \quad \mathscr{U} = \mathbb{R}^d, \quad \mathcal{U} = \mathcal{D}(\mathbb{R}^d;\mathbb{R}^d), \\ |||\cdot||| = \begin{cases} \|\cdot\|_{L^q_\mu} & \text{if} \quad q \in [1,\infty), \\ \|\cdot\|_{\infty,\Omega} & \text{if} \quad q = \infty, \end{cases} \\ V = \mathcal{S}^{d\times d}, \quad Au = e(u) := \frac{1}{2}\Big(\nabla u + (\nabla u)^\top\Big). \end{cases}$$

Verification of stability condition ($\mathrm{H}_{\mathrm{stab}}$) is in full analogy to the case of $A = \nabla$, cf. Example 2.49. Similarly as in the last subsection, for the operator e we reserve notation as below:

$$A_\mu =: e_\mu, \qquad \mathrm{dom}(A_\mu) =: \mathscr{D}^{1,q}_\mu \qquad \Pi_\mu =: \Pi^e_\mu, \qquad M_\mu =: M^e_\mu.$$

The letter $\mathscr{D}$ ought to stand for *deformation*. Due to the lack of Korn's inequality for general measures the relation between the tangential spaces T_μ and M^e_μ may be *a priori* established only as an inclusion:

Proposition 2.66. *For any compactly supported measure $\mu \in \mathcal{M}_+(\mathbb{R}^d)$ and any exponent $q \in [1,\infty]$ there holds*

$$\big(W^{1,q}_\mu\big)^d \subset \mathscr{D}^{1,q}_\mu, \qquad M^e_\mu \subset T_\mu \overset{s}{\otimes} T_\mu := \left\{ \sum_{i=1}^d a_i\, \tau_i \otimes \tau_i \,:\, a_i \in \mathbb{R},\ \tau_i \in T_\mu \right\} \quad \mu\text{-a.e.}$$

Moreover, if $u \in (W^{1,q}_\mu)^d$ then

$$e_\mu(u) = \Pi^e_\mu\left(\frac{1}{2}\Big(\nabla_\mu u + (\nabla_\mu u)^\top\Big)\right). \tag{2.25}$$

Proof. For a fixed $u \in (W^{1,q}_\mu)^d$ we may choose a sequence $\{u_n\} \subset \mathcal{U}$ and $\epsilon \in L^q_\mu(\mathbb{R}^d;\mathbb{R}^{d\times d})$ such that $u_n \to u$ in $\overline{\mathcal{U}}$ and $\nabla u_n \to \epsilon$ (weakly-* if $q = \infty$), therefore $\nabla_\mu u = P_\mu \epsilon$. Then it is straightforward that $e(u_n) = \frac{1}{2}\big(\nabla u_n + (\nabla u_n)^\top\big) \to \frac{1}{2}\big(\epsilon + \epsilon^\top\big)$ in $L^q_\mu(\mathbb{R}^d;\mathcal{S}^{d\times d})$. As a consequence $u \in \mathscr{D}^{1,q}_\mu$ and $e_\mu(u) = \Pi^e_\mu\big(\frac{1}{2}(\epsilon + \epsilon^\top)\big)$. Consequently the normal spaces $\mathcal{N} = (\mathcal{N}^\nabla)^d$ and $\mathcal{N} = \mathcal{N}^e$ computed for $A = \nabla$ and $A = e$, respectively, satisfy the relation

$$\left\{ \frac{1}{2}\big(\epsilon + \epsilon^\top\big) : \epsilon \in \big(\mathcal{N}^\nabla\big)^d \right\} \subset \mathcal{N}^e.$$

which for μ-a.e. x implies that $\big\{\frac{1}{2}\big(\xi + \xi^\top\big) : \xi \in \big(N^\nabla_\mu(x)\big)^d\big\} \subset N^e_\mu(x)$. Let as now assume that $\sigma \in M^e_\mu(x)$; as an element of $\mathcal{S}^{d\times d}$ the matrix σ may be written as $\sum_{i=1}^d \lambda_i\, \tau_i \otimes \tau_i$ where $\lambda_i \in \mathbb{R}$, $\tau_i \in S^{d-1}$ and $\langle \tau_i, \tau_j\rangle = 0$ when $i \neq j$. There must hold $\sigma \perp \xi$ for any

$\xi \in N^e_\mu(x)$ and thus, in particular for $\xi = \sum_{j=1}^d \frac{1}{2}\big(\tau_j \otimes v_j + v_j \otimes \tau_j\big)$ where v_j are any elements in $N^\nabla_\mu(x)$. Hence there must hold

$$0 = \langle \xi, \sigma \rangle = \sum_{i,j=1}^d \lambda_i \, \langle \tau_i, \tau_j \rangle \langle \tau_i, v_j \rangle = \sum_{i=1}^d \lambda_i \, \langle \tau_i, v_i \rangle.$$

Since v_i may be chosen arbitrarily from $N^\nabla_\mu(x)$ we see that this implies $\tau_i \in M^\nabla_\mu(x) = T_\mu(x)$ whenever $\lambda_i \neq 0$ thus furnishing the second asserted inclusion. This inclusion implies that $\Pi^e_\mu\big(P_\mu\epsilon + (P_\mu\epsilon)^\top\big) = \Pi^e_\mu(\epsilon + \epsilon^\top)$, which furnishes (2.25). □

The converse inclusion may be established for multi-junction measures μ:

Proposition 2.67. *For a multi-junction measure μ the space of tangent matrices reads*

$$M^e_\mu = T_\mu \overset{s}{\otimes} T_\mu = T_{\mu_i} \overset{s}{\otimes} T_{\mu_i} = T_{S_i} \overset{s}{\otimes} T_{S_i} \qquad \mu_i\text{-a.e.} \tag{2.26}$$

where T_{S_i} denotes the space tangent to manifold S_i in the classical sense. Moreover for $\xi \in \mathcal{S}^{d\times d}$

$$\Pi^e_\mu \xi = P_\mu \, \xi \, P_\mu = P_{\mu_i} \, \xi \, P_{\mu_i} = P_{S_i} \, \xi \, P_{S_i} \qquad \mu_i\text{-a.e.} \tag{2.27}$$

Proof. We shall first show that $M^e_{\mu_i} = T_{\mu_i} \overset{s}{\otimes} T_{\mu_i}$ for each i. By combining Propositions 2.59 and 2.66 we obtain $M^e_{\mu_i} \subset T_{\mu_i} \overset{s}{\otimes} T_{\mu_i} = T_{S_i} \overset{s}{\otimes} T_{S_i}$. We can see that any function $\sigma \in C^2(S_i; T_{S_i} \overset{s}{\otimes} T_{S_i})$ that is zero on ∂S_i satisfies $\int_{S_i} \langle e(\varphi), \sigma \rangle \, d\mathcal{H}^{k_i} = 0$ by virtue of integration by parts formula on manifolds. By Proposition 2.48 and characterization of M_μ in Proposition 2.50 we infer the converse inclusion $M^e_\mu(x) \supset T_{S_i} \overset{s}{\otimes} T_{S_i} = T_{\mu_i} \overset{s}{\otimes} T_{\mu_i}$ $\mathcal{H}^{k_i}$-a.e.

Since $\mu_i \leq \mu$ it is straightforward to check that $M^e_{\mu_i} \subset M^e_\mu$ μ_i-a.e. while, once more by Proposition 2.66, we obtain $M^e_\mu \subset T_\mu \overset{s}{\otimes} T_\mu$ while $T_\mu \overset{s}{\otimes} T_\mu = T_{\mu_i} \overset{s}{\otimes} T_{\mu_i}$ μ_i-a.e. thanks to Proposition 2.59. Ultimately, μ_i-a.e.

$$T_{S_i} \overset{s}{\otimes} T_{S_i} = T_{\mu_i} \overset{s}{\otimes} T_{\mu_i} = M^e_{\mu_i} \subset M^e_\mu \subset T_\mu \overset{s}{\otimes} T_\mu = T_{\mu_i} \overset{s}{\otimes} T_{\mu_i} = T_{S_i} \overset{s}{\otimes} T_{S_i}$$

and (2.26) is established. It is easy to verify that the orthogonal projection operator Π^e_μ is characterized via (2.27). □

2.5.4. The μ-tangential Hessian operator for $q = \infty$

The last framework that we shall herein specify concerns the second-order Hessian operator $A = \nabla^2$. We must, however, confine ourselves to the case of $q = \infty$ (cf. Remark 2.68 below): for bounded domain Ω satisfying $\operatorname{spt}\mu \subset \overline{\Omega}$ we take

$$\begin{cases} U = \mathbb{R}, \quad \mathscr{U} = \mathbb{R}^d, \quad \mathcal{U} = \mathcal{D}(\mathbb{R}^d), \\ |||u||| = \|u\|_{\infty,\Omega} + \|\nabla u\|_{\infty,\Omega} \\ V = \mathcal{S}^{d\times d}, \quad A = \nabla^2. \end{cases} \tag{2.28}$$

To check the stability condition ($\mathrm{H}_{\mathrm{stab}}$), for $\epsilon \in \mathcal{N}$ we take a sequence $\{u_n\} \subset \mathcal{D}(\mathbb{R}^d)$ such that $\|u_n\|_{\infty,\Omega} \to 0$, $\|\nabla u_n\|_{\infty,\Omega} \to 0$ and $\nabla^2 u_n \overset{*}{\rightharpoonup} \epsilon$ in $L^\infty_\mu(\mathbb{R}^d; \mathcal{S}^{d\times d})$ (such sequence exists by definition). Then, for any cut-off function $\psi \in \mathcal{D}(\mathbb{R}^d)$

$$\nabla(\psi u_n) = u_n \nabla\psi + \psi \, \nabla u_n, \qquad \nabla^2(\psi u_n) = u_n \nabla^2\psi + 2 \, \nabla\psi \overset{s}{\otimes} \nabla u_n + \psi \, \nabla^2 u_n \tag{2.29}$$

and we readily see that $|||\psi u_n||| \to 0$ and $\nabla^2(\psi u_n) \overset{*}{\rightharpoonup} \psi\epsilon$ therefore $\psi\epsilon \in \mathcal{N}$ validating ($\mathrm{H}_{\mathrm{stab}}$).

Remark 2.68. For $q < \infty$ the choice of the norm $|||\cdot|||$ is not obvious in light of condition ($\mathrm{H}_{\mathrm{stab}}$). With $|||\cdot|||$ defined as in (2.28) the stability condition is clearly met for any q, however this particular norm deprives us of coerciveness when working with variational frameworks defined on measure, e.g. for elasticity of Kirchhoff plates. This issue was addressed in [36] where the second-order operator A_μ admitted two variables: function u and independent *Cosserat vector field* b that point-wise lies in the normal space N_μ.

For consistency we propose notation for the second-order μ-tangential calculus presently discussed:

$$A_\mu =: \nabla^2_\mu, \qquad \mathrm{dom}(A_\mu) =: W^{2,\infty}_\mu, \qquad \Pi_\mu =: \Pi^{\nabla^2}_\mu, \qquad M_\mu =: M^{\nabla^2}_\mu$$

Once more the space of tangent matrices $M^{\nabla^2}_\mu$ can be expressed through T_μ provided μ is assume to be a multi-junction measure. A considerable difference with respect to first-order elasticity discussed in Section 2.5.3 may be observed:

Proposition 2.69. *If μ is a multi-junction measure then for every $\xi \in \mathcal{S}^{d\times d}$ there μ-a.e. holds*

$$\mathcal{S}^{d\times d} = M^{\nabla^2}_\mu \oplus \left(T^\perp_\mu \overset{s}{\otimes} T^\perp_\mu\right), \qquad \Pi^{\nabla^2}_\mu \xi = P_\mu\,\xi\,P_\mu + P^\perp_\mu\,\xi\,P_\mu + P_\mu\,\xi\,P^\perp_\mu,$$

where $\oplus$ stand for the direct sum and $P^\perp_\mu = \mathrm{I} - P_\mu$ is the orthogonal projection onto $T^\perp_\mu = N^\nabla_\mu$.

Proof. See [36, Proposition 3.9] and Section 5 of [37]. □

2.6. Conic programming

In the present section we review the basic facts on duality in finite dimensional conic programming that shall be the main numerical tool employed in this dissertation. Conic programming problems (or, shortly, *conic programs*) can be viewed as a generalization of linear programming problems through adding the so called conic constraints. Below we draw upon Section 2 of book [14] where the reader is referred for details.

For a closed convex cone $\mathcal{K}$ in $\mathbb{R}^d$ by the *dual cone* $\mathcal{K}^*$ we understand

$$\mathcal{K}^* := \Big\{ z^* \in \mathbb{R}^d \,:\, \langle z, z^*\rangle \geq 0 \quad \forall\, z \in \mathcal{K} \Big\}. \tag{2.30}$$

For positive integers N, M let there be given vectors $\mathbf{c} \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^M$ and a matrix $\mathbf{A} \in \mathbb{R}^{M\times N}$. Next, we consider a family of L cones: $\mathcal{K}_i \subset \mathbb{R}^{d(i)}$ for $i \in \{1, \ldots, L\}$ where $d(i)$ is a positive integer; in addition for each i we assume a matrix $\mathbf{Q}_i \in \mathbb{R}^{d(i)\times N}$.

By a conic program (CP) we understand the following problem:

$$\inf_{\mathbf{x}\in\mathbb{R}^N} \left\{ \langle \mathbf{c}, \mathbf{x}\rangle \,:\quad \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \begin{matrix} \mathbf{Q}_1\mathbf{x} \in \mathcal{K}_1 \\ \vdots \\ \mathbf{Q}_L\mathbf{x} \in \mathcal{K}_L \end{matrix} \right\} \tag{CP}$$

whereas the dual problem (CP*) reads

$$\sup_{\substack{\mathbf{y}\in\mathbb{R}^M \\ \boldsymbol{\eta}_1\in\mathbb{R}^{d(1)},\,\ldots,\,\boldsymbol{\eta}_L\in\mathbb{R}^{d(L)}}} \left\{ \langle \mathbf{b},\mathbf{y}\rangle \;:\; \begin{array}{c} \mathbf{A}^\top\mathbf{y}+\sum_{i=1}^{L}\mathbf{Q}_i^\top\boldsymbol{\eta}_i=\mathbf{c}, \\ \\ \end{array} \begin{array}{c} \boldsymbol{\eta}_1\in\mathcal{K}_1^* \\ \vdots \\ \boldsymbol{\eta}_L\in\mathcal{K}_L^* \end{array} \right\} \tag{CP*}$$

In order to give rigorous duality results we require that the conic program meets the *constraint's qualification*:

Definition 2.70. Assume that:

(1) rows of matrix $\mathbf{A}$ are linearly independent, i.e. $\operatorname{Im}\mathbf{A}=\mathbb{R}^M$,
(2) $\operatorname{Ker}\mathbf{A}\cap\left(\bigcap_{i=1}^{L}\operatorname{Ker}\mathbf{Q}_i\right)$ is the trivial subspace of $\mathbb{R}^N$.

Then, we say that (CP) is *strictly feasible* if there exists a feasible vector $\mathbf{x}\in\mathbb{R}^N$ such that $\mathbf{Q}_i\mathbf{x}\in\operatorname{int}\mathcal{K}_i$ for each $i\in\{1,\ldots,L\}$.

Analogously, we say that (CP*) is strictly feasible if there exist feasible vectors $\mathbf{y}\in\mathbb{R}^M,\boldsymbol{\eta}_1\in\mathbb{R}^{d(1)},\;\ldots\;,\boldsymbol{\eta}_L\in\mathbb{R}^{d(L)}$ such that $\boldsymbol{\eta}_i\in\operatorname{int}\mathcal{K}_i^*$ for each $i\in\{1,\ldots,L\}$.

Theorem 2.71. *The following statements hold:*

(i) If one of the problems (CP) *or* (CP*) *is strictly feasible and bounded, then the duality gap vanishes:*

$$\sup\mathrm{CP}^*=\inf\mathrm{CP}<\infty.$$

(ii) If (CP) *is strictly feasible and bounded then* (CP*) *admits a solution.*
(iii) If (CP*) *is strictly feasible and bounded then* (CP) *admits a solution.*
(iv) Assuming that (i) holds, feasible vectors $\mathbf{x}\in\mathbb{R}^N$ and $\mathbf{y}\in\mathbb{R}^M,\boldsymbol{\eta}_1\in\mathbb{R}^{d(1)},\;\ldots\;,\boldsymbol{\eta}_L\in\mathbb{R}^{d(L)}$ are solutions of (CP) *and* (CP*)*, respectively, if and only if the following complementary slackness conditions are met:*

$$\langle\,\mathbf{Q}_i\mathbf{x}\,,\,\boldsymbol{\eta}_i\,\rangle_{\mathbb{R}^{d(i)}}=0\qquad\forall\,i\in\{1,\ldots,L\}.$$

Proof. See [14, Section 2.5]. □

In the available software conic programs are usually implemented for several simple classes of cones $\mathcal{K}$. Perhaps the most classical type of such programs are *conic quadratic programming problems* where only cones of revolution are used; in [9] the authors propose to use two cones only: the *quadratic cone*

$$\mathcal{K}^d_{\mathrm{quad}}:=\left\{\mathbf{z}\in\mathbb{R}^d\;:\;\mathbf{z}(1)\geq\sqrt{\big(\mathbf{z}(2)\big)^2+\ldots+\big(\mathbf{z}(d)\big)^2}\right\} \tag{2.31}$$

and, for $d\geq 3$, the *rotated quadratic cone*

$$\mathcal{K}^d_{\mathrm{rot,quad}}:=\left\{\mathbf{z}\in\mathbb{R}^d\;:\;\mathbf{z}(1),\mathbf{z}(2)\geq 0,\;\;2\,\mathbf{z}(1)\,\mathbf{z}(2)\geq\big(\mathbf{z}(3)\big)^2+\ldots+\big(\mathbf{z}(d)\big)^2\right\}. \tag{2.32}$$

The two cones are self-dual, i.e.

$$\big(\mathcal{K}^d_{\mathrm{quad}}\big)^*=\mathcal{K}^d_{\mathrm{quad}},\qquad\big(\mathcal{K}^d_{\mathrm{rot,quad}}\big)^*=\mathcal{K}^d_{\mathrm{rot,quad}}.$$

Chapter 3

Including the choice of anisotropy in the optimal design

This thesis covers three topics, each concerns a different optimal design formulation:

- the class of Free Material Design problems (Chapter 4),
- the optimal membrane problem (Chapter 5),
- the optimal form-finding problem (Chapter 6).

In each of those formulations, at one point or another, we shall be seeking the field of Hooke (stiffness) tensor that minimizes compliance of an elastic body or structure. The goal of the present chapter is to prepare a unified mathematical framework that facilitates this type of design. Although elasticity is our primary concern, digressions on design of heat conductors will be made hence the abstract setting of this chapter: generic finite dimensional spaces V, W will be used and we will mainly speak of anisotropy parameters rather than of Hooke tensors.

3.1. The energy potential parametrized by the anisotropy

In this chapter V will stand for arbitrary finite dimensional space endowed with a scalar product $\langle \cdot, \cdot \rangle$. Classically, elements of dual space V^* will be identified with vectors in V. In the remaining chapters of this thesis two variants of V will be considered depending on the physical context:

1. In case of d-dimensional elasticity $V = \mathcal{S}^{d\times d}$ being the space of symmetric $d \times d$ matrices: the elements $\xi \in V$ represent the strain tensors while $\sigma \in V^* \equiv \mathcal{S}^{d\times d}$ the stress tensors; the duality pairing is the Euclidean scalar product $\langle \xi, \sigma \rangle = \sum_{i,j=1}^{d} \xi_{ij}\sigma_{ij}$.
2. In case of d-dimensional conductivity $V = \mathbb{R}^d$: the elements $\xi \in V$ should be interpreted as minus gradients of temperature field whereas $\sigma \in V^* \equiv \mathbb{R}^d$ represents the heat flow; the duality pairing is the standard Euclidean product.

Next we choose another finite dimensional space W and then

$$\text{we fix any closed convex cone } \mathcal{K} \subset W. \tag{3.1}$$

The elements $H \in \mathcal{K}$ will parametrize the constitutive law of material: essentially H shall determine the material's anisotropy. We give some basic examples that we will consider, once again depending on the physical context:

Example 3.1. In case of elasticity, where $V = \mathcal{S}^{d\times d}$, it is natural to take $W = \mathscr{L}(\mathcal{S}^{d\times d})$ being the space of symmetric linear operators from $\mathcal{S}^{d\times d}$ to $\mathcal{S}^{d\times d}$. Then, we choose $\mathcal{K}$ as any closed convex subcone of the cone $\mathscr{L}_+(\mathcal{S}^{d\times d})$ consisting only of the positive semi-definite operators. The cone $\mathscr{L}_+(\mathcal{S}^{d\times d})$ itself is isomorphic to the cone of 4th order *elasticity* or *Hooke tensors* (hence the symbol $H \in \mathcal{K}$) that establish the constitutive law in classical elasticity. The basic possible choice would be

$$\mathcal{K} = \mathcal{K}_{ani} := \mathscr{L}_+(\mathcal{S}^{d\times d}),$$

admitting arbitrary anisotropic Hooke law, or for instance

$$\mathcal{K} = \mathcal{K}_{iso} := \left\{ d\mathscr{K}\left(\frac{1}{d}\,\mathrm{I}\otimes\mathrm{I}\right) + 2\mathscr{G}\left(\mathrm{Id} - \frac{1}{d}\,\mathrm{I}\otimes\mathrm{I}\right) \;:\; \mathscr{K},\mathscr{G} \geq 0 \right\} \subset \mathscr{L}_+(\mathcal{S}^{d\times d}), \quad (3.2)$$

which enforces isotropy of the material; above $\mathrm{I} \in \mathcal{S}^{d\times d}$ is the identity matrix while $\mathrm{Id} \in \mathscr{L}(\mathcal{S}^{d\times d})$ is the identity operator; $\mathscr{K}$ and $\mathscr{G}$ are the *bulk* and, respectively, *shear moduli*. △

Example 3.2. Similarly, for the conductivity context where $V = \mathbb{R}^d$ naturally $W = \mathscr{L}(\mathbb{R}^d)$ which is isomorphic to $\mathcal{S}^{d\times d}$ and thus we shall write $W = \mathcal{S}^{d\times d}$ instead. Then, the choice

$$\mathcal{K} = \mathcal{K}_{ani,c} := \mathcal{S}^{d\times d}_+$$

represents the family of anisotropic conductivity tensors, where $\mathcal{S}^{d\times d}_+$ is the cone of positive semi-definite symmetric matrices. Imposing isotropy of the conductor reduces the set to a half-line

$$\mathcal{K} = \mathcal{K}_{iso,c} := \left\{ a\,\mathrm{I} \;:\; a \geq 0 \right\} \subset \mathcal{S}^{d\times d}_+.$$

△

In case of elasticity the variety of choice of $\mathcal{K}$ being a closed convex cone in $\mathscr{L}_+(\mathcal{S}^{d\times d})$ is certainly wide, which we shall bring into focus in Section 3.3.

Throughout the whole work j will stand for a two-argument energy fnction:

$$j : \mathcal{K} \times V \to \mathbb{R},$$

while, for a fixed exponent $p \in (1,\infty)$, we shall assume that:

- for each $H \in \mathcal{K}$:
  (A1) the function $j(H,\,\cdot\,)$ is real valued, non-negative and convex on V;
  (A2) the function $j(H,\,\cdot\,)$ is positively p-homogeneous on V;
- whilst for each $\xi \in V$:
  (A3) the function $j(\,\cdot\,,\xi)$ is concave and upper semi-continuous on $\mathcal{K}$;
  (A4) the function $j(\,\cdot\,,\xi)$ is positively one-homogeneous on $\mathcal{K}$.

For each $H \in \mathcal{K}$ the function $j(H,\,\cdot\,) : V \to \mathbb{R}$ becomes an energy potential: we say that the pair $\xi, \sigma \in V$ satisfies the material's constitutive law whenever

$$\sigma \in \partial j(H,\xi), \quad (3.3)$$

where for each $H \in \mathcal{K}$ we shall henceforward agree that $\partial j(H,\,\cdot\,) \subset V^* \equiv V$ denotes the subdifferential of the convex function $j(H,\,\cdot\,)$, i.e. the subdifferential is taken with

respect to the second argument of j. Similarly we will understand the convex conjugate $j^*(H,\,\cdot\,) : V^* \equiv V \to \overline{\mathbb{R}} = [-\infty,\infty]$ that is again computed with respect to the second argument, namely:

$$j^*(H,\sigma) = \sup_{\zeta\in V}\Big\{\langle\zeta,\sigma\rangle - j(H,\zeta)\Big\}. \tag{3.4}$$

As a result, thanks to (A1),(A2), it is well established (cf. Proposition 2.14) that for a given $H \in \mathcal{K}$ the constitutive law (3.3) is equivalent to the *extremality equation*:

$$\sigma \in \partial j(H,\xi) \qquad \Leftrightarrow \qquad \langle\xi,\sigma\rangle = j(H,\xi) + j^*(H,\sigma), \tag{3.5}$$

while the inequality $\leq$ on the RHS, known as Fenchel's inequality, holds always. Owing to p-homogeneity of $j(H,\,\cdot\,)$ we obtain:

Proposition 3.3. *For any $\xi,\sigma \in V$ a rule of repartition of energy holds:*

$$\sigma \in \partial j(H,\xi) \qquad \Leftrightarrow \qquad \begin{cases} \langle\xi,\sigma\rangle = p\ j(H,\xi), \\ \langle\xi,\sigma\rangle = p'\, j^*(H,\sigma). \end{cases} \tag{3.6}$$

Proof. Implication $\Leftarrow$ easily comes from implication $\Leftarrow$ in (3.5) and the fact that $1/p + 1/p' = 1$. Conversely, assuming that $\sigma \in \partial j(H,\xi)$ again by (3.5) we get $\langle\xi,\sigma\rangle - j(H,\xi) = \sup_{\zeta\in V}\big\{\langle\zeta,\sigma\rangle - j(H,\zeta)\big\} \geq \sup_{t\geq 0}\big\{\langle t\,\xi,\sigma\rangle - j(H,t\,\xi)\big\} = \sup_{t\geq 0}\big\{t\langle\xi,\sigma\rangle - t^p j(H,\xi)\big\}$ and the Euler-Lagrange equation with respect to t reads $\langle\xi,\sigma\rangle = p\,t^{p-1}j(H,\xi)$. However, by the chain of inequalities we know that $t = 1$ is the solution and this gives $\langle\xi,\sigma\rangle = p\,j(H,\xi)$, while the second equality is by RHS of (3.5). □

Example 3.4. Once $\mathcal{K}$ is a subcone of $\mathscr{L}_+(V)$, being the cone of symmetric and positive semi-definite operators from V to V (cf. Examples 3.1,3.2), the simplest choice of the function j that satisfies assumptions (A1)-(A4) would be:

$$j(H,\xi) = \frac{1}{2}\langle H\xi,\xi\rangle$$

for which the associated constitutive law (3.3) is a linear equation: $\sigma = H\xi$. △

Remark 3.5. We note that in order for the assumptions (A1)-(A4) to hold the set $\mathcal{K}$ (or more precisely $\mathcal{K} \cap \bigcup_{\xi\in V}\operatorname{spt}\big(j(\,\cdot\,,\xi)\big)$) has to be contained in a half-space of W. Indeed, otherwise there would exist $\xi_0 \in V$ and a straight line $\big\{tH_0 : t \in \mathbb{R}\big\} \subset \mathcal{K}$ such that $j(H_0,\xi_0) > 0$. But the function $t \mapsto j(tH_0,\xi_0)$ must be both concave, non-negative and positive 1-homogeneous which is impossible. If, on the other hand, $\mathcal{K} \subset \mathscr{L}_+(V)$ a valid function j may be proposed so that $j(\,\cdot\,,\xi)$ is non-trivially concave, i.e. it is not linear. An important function j of this type will be given in Example 3.24, Section 3.3.

In convex analysis a convex function restricted to a convex subset of a linear space can be equivalently treated as a function defined on the whole space if it is extended by $+\infty$. Since the real function $j : \mathcal{K}\times V \to \mathbb{R}$ is concave with respect to first variable H we can by analogy speak of an extended real function $j : W \times V \to \overline{\mathbb{R}} = [-\infty,\infty]$ such that $j(H,\xi) = -\infty$ for any $\xi \in V$ and any $H \in W\backslash\mathcal{K}$. Since the set $\mathcal{K}$ is closed the function $j(\,\cdot\,,\xi)$ remains upper semi-continuous for every ξ.

Since for each $H \in \mathcal{K}$ the function $j(H, \cdot\,)$ is convex and finite on the whole finite dimensional space V the classical result (cf. Corollary 2.7) guarantees that j is continuous in ξ. Independently, we have upper semi-continuity of j in the first variable H owing to assumption (A3). The next result establishes upper semi-continuity of the function j extended by $-\infty$ jointly in variables H and ξ:

Proposition 3.6. *The function j is upper semi-continuous on the product $W \times V$.*

Proof. We fix a pair $(\breve{H}, \breve{\xi}) \in W \times V$. We may assume that $\breve{H} \in \mathcal{K}$ since $j = -\infty$ in $(W \setminus \mathcal{K}) \times V$ which is an open set. Let us take any ball $\mathscr{U} \subset W$ centred at $\breve{H}$ and introduce a compact set $K = \overline{\mathscr{U}} \cap \mathcal{K}$. We observe that for any fixed $\xi \in V$ the set $\{j(H, \xi) : H \in K\}$ is bounded in $\mathbb{R}$. The zero lower bound follows from non-negativity of $j(\,\cdot\,, \xi)|_{\mathcal{K}}$, whereas $j(\,\cdot\,, \xi)$ achieves its finite maximum on the compact set K as it is real valued, concave and upper semi-continuous on $\mathcal{K} \supset K$. According to [140, Theorem 10.6] the shown point-wise boundedness combined with convexity of every $j(H, \cdot\,)$ imply that the family of functions $\{j(H, \cdot\,) : H \in K\}$ is equi-continuous on any bounded subset of V. Upon fixing $\varepsilon > 0$ we may thus choose $\delta_1 > 0$ such that

$$\left| j(H, \xi) - j(H, \breve{\xi}) \right| < \frac{\varepsilon}{2} \qquad \forall\, \xi \in B(\breve{\xi}, \delta_1) \subset V, \quad \forall\, H \in K \subset \mathcal{K},$$

where it must be stressed that K does not depend on ε. Due to the upper semi-continuity of $j(\,\cdot\,, \breve{\xi})$ we can also choose $\delta_2 > 0$ for which $B(\breve{H}, \delta_2) \subset \mathscr{U}$ and

$$j(H, \breve{\xi}) < j(\breve{H}, \breve{\xi}) + \frac{\varepsilon}{2} \qquad \forall\, H \in B(\breve{H}, \delta_2).$$

For any pair $(H, \xi) \in \big(B(\breve{H}, \delta_2) \cap \mathcal{K}\big) \times B(\breve{\xi}, \delta_1)$ we therefore obtain

$$j(H, \xi) = j(H, \breve{\xi}) + \big(j(H, \xi) - j(H, \breve{\xi})\big) < j(\breve{H}, \breve{\xi}) + \frac{\varepsilon}{2} + \frac{\varepsilon}{2},$$

which proves that j is upper semi-continuous on $\mathcal{K} \times V$ being a convex closed subset of $W \times V$. Extending j by $-\infty$ guarantees it is upper semi-continuous on $W \times V$. □

We now turn to study the function $j^* : \mathcal{K} \times V \to \overline{\mathbb{R}}$ which can be similarly extended to $j^* : W \times V \to \overline{\mathbb{R}}$: since $j = -\infty$ on $(H, \xi) \in (W \setminus \mathcal{K}) \times V$ it is straightforward to compute that $j^* = \infty$ on $(W \setminus \mathcal{K}) \times V$. It is well-established that $j^*(H, \cdot\,)$ as a convex conjugate is convex and lower semi-continuous on V, although we can infer more:

Proposition 3.7. *The function j^* is convex and lower semi-continuous on $W \times V$, i.e. jointly in arguments H and σ. Moreover, for each $H \in \mathcal{K}$ the function $j^*(H, \cdot\,) : V \to \overline{\mathbb{R}}$ is proper, non-negative and postitively p'-homogeneous while*

$$j^*(tH, \sigma) = \frac{1}{t^{\,1/(p-1)}}\, j^*(H, \sigma) \qquad \forall\, t > 0. \tag{3.7}$$

Proof. Since $j(\,\cdot\,, \zeta)$ is concave and upper semi-continuous for every $\zeta \in V$ the mapping $(H, \sigma) \mapsto \langle \zeta, \sigma \rangle - j(H, \zeta)$ is convex and lower semi-continuous jointly in H and σ. As a result the function $j^* : W \times V \to \overline{\mathbb{R}}$ is also jointly convex and lower semi-continuous as a point-wise supremum with respect to ζ, cf. the definition (3.4). Next, once $H \in \mathcal{K}$ we have $j(H, 0) = 0$ guaranteeing that $j^*(H, \cdot\,)$ is non-negative, while non-negativity of $j(H, \cdot\,)$ implies that $j^*(H, 0) = 0$ hence $j^*(H, \cdot\,)$ is proper. For the positive p'-homogeneity of

$j^*(H, \cdot)$ we refer to [140, Corollary 15.3.1.]. The formula (3.7) follows from the known equality $(tf)^*(\cdot) = tf^*(\cdot/t)$ (see e.g. [80, Chapter I, Eq. (4.7)]) and positive 1-homogeneity of $j(\cdot, \xi)$:

$$j^*(tH, \sigma) = \big(j(tH, \cdot)\big)^*(\sigma) = \big(t\, j(H, \cdot)\big)^*(\sigma) = t\,\big(j(H, \cdot)\big)^*(\sigma/t)$$
$$= t\, j^*(H, \sigma/t) = \frac{t}{t^{p'}}\, j^*(H, \sigma).$$

□

Remark 3.8. Although j is a real valued function on $\mathcal{K} \times V$ we find that j^*, in general, may admit infinite values: take for instance $H \in \mathcal{K} \in \mathscr{L}_+(V)$ that is a singular operator and $j(H, \xi) = \frac{1}{2}\langle H\,\xi, \xi\rangle$, then $j^*(H, \sigma) = \infty$ for any $\sigma \not\perp \operatorname{Ker} H$. As it shall appear for important problems considered in this thesis for certain $\sigma_0 \in V$ the function $j^*(\cdot, \sigma_0)$ will be constantly equal to ∞ and thus non-proper. In order to avoid this a certain ellipticity must be imposed on j, cf. assumption $(\mathrm{A}_{\mathrm{ellip}})$ in Section 3.2 below.

3.2. Designing anisotropy at a point

We begin by specifying requirements for the unit cost function $c : \mathcal{K} \subset W \to \mathbb{R}_+$ that establishes the global constraint on the constitutive field. Throughout the text

we assume that the cost function c is a norm on W restricted to $\mathcal{K}$.

The symmetry property of norms is not crucial here, i.e. we could take a function that is positive 1-homogeneous instead of 1-homogeneous. However, this matter becomes immaterial due to Remark 3.5.

Example 3.9. For the space $W = \mathscr{L}(V)$ (cf. Examples 3.1, 3.2) it is natural to choose $c = c(H)$ as one of the following norms explicitly given in terms of eigenvalues $\lambda_i(H)$:

$$c_q(H) = \left(\sum_{i=1}^{N} |\lambda_i(H)|^q\right)^{\frac{1}{q}}, \qquad c_\infty(H) = \max_{i\in\{1,\dots,N\}} |\lambda_i(H)|,$$

where $q \in [1, \infty)$ and $N = \dim(V)$. If we assume that $\mathcal{K} \subset \mathscr{L}_+(V)$ then for each $H \in \mathcal{K}$ the c_1 norm attains a special form:

$$c_1(H) = \sum_{i=1}^{N} \lambda_i(H) = \operatorname{Tr} H$$

i.e. $c_1 = \operatorname{Tr}$ is linear on $\mathcal{K}$. This trace cost shall be the only one considered in examples in this thesis for it is not only the most convenient mathematically, but also the most natural in terms of physics, cf. the pioneering work on the Free Material Design problem [136]. △

In Chapters 4, 5, 6 we will be optimizing the anisotropy field being a $\mathcal{K}$-valued measure λ that is subject to the constraint on the total cost: $\int c(\lambda) \le C_0$. Such measure may be decomposed to $\lambda = \mathscr{H}\mu$ where $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\mathscr{H} \in L^1_\mu(\overline{\Omega}; \mathcal{K})$. As it will unravel in subsequent chapters it will be in our interest to maximize the strain energy $\int j(\mathscr{H}, \epsilon)\, d\mu$ and (more importantly) minimize the complementary stress energy $\int j^*(\mathscr{H}, \sigma)\, d\mu$. If the

normalization $c(\mathscr{H}) = 1$ μ-a.e. is imposed then the constraint takes the form $\int d\mu \le C_0$ and optimization of the integral with respect to the anisotropy becomes local.

The above reasoning shall be made rigorous in Section 3.4 and then in the remaining chapters, nevertheless it already motivates us to look at the finite dimensional program

$$\bar{j}(\xi) := \max_{H \in \mathcal{K}_1} j(H, \xi), \tag{3.8}$$

where the maximization is carried out over a c-normalization of the set $\mathcal{K}$:

$$\mathcal{K}_1 := \Big\{ H \in \mathcal{K} \ : \ c(H) \le 1 \Big\}.$$

Since c is a norm the set $\mathcal{K}_1$ is a convex compact subset of W. Owing to assumption (A3) we immediately infer that problem (3.8) has a solution. It will be convenient to give a symbol to the set of solutions:

$$\bar{\mathcal{K}}_1(\xi) := \Big\{ H \in \mathcal{K}_1 : j(H, \xi) = \bar{j}(\xi) \Big\}$$

which for any $\xi \in V$ is another convex compact non-empty set in W.

Remark 3.10. The short over-bar $\bar{\cdot}$ will be consistently used in the sequel to mark maximization with respect to anisotropy H and should not be confused with long over-bar $\overline{\cdot}$ denoting e.g. the closure of a set. Similarly, in due course, we shall use the under-bar $\underline{\cdot}$ to stress minimization over the set $\mathcal{K}_1$.

The function $\bar{j} : V \to \mathbb{R}$ represents energy that is maximal with respect to anisotropy parameter $H \in \mathcal{K}$ of a unit c-cost. It inherits the properties of the functions $j(H, \cdot\,)$:

Proposition 3.11. *The function $\bar{j} : V \to \mathbb{R}$ is real valued, non-negative, convex, continuous and positively p-homogeneous. As a consequence there exists a unique closed gauge (function that is non-negative, convex, positive 1-homogeneous and lower semi-continuous) $\rho : V \to \mathbb{R}_+$ such that*

$$\bar{j}(\xi) = \frac{1}{p} \big(\rho(\xi)\big)^p. \tag{3.9}$$

The gauge ρ is real valued and continuous. The convex conjugate $\bar{j}^* := (\bar{j})^* : V^* \equiv V \to \overline{\mathbb{R}}$ can be expressed by the formula

$$\bar{j}^*(\sigma) = \frac{1}{p'} \big(\rho^0(\sigma)\big)^{p'}, \qquad \frac{1}{p} + \frac{1}{p'} = 1, \tag{3.10}$$

where $\rho^0 : V^* \equiv V \to \overline{\mathbb{R}}$ is the closed gauge polar to ρ, i.e.

$$\rho^0(\sigma) = \sup_{\zeta \in V} \Big\{ \langle \zeta, \sigma \rangle \ : \ \rho(\zeta) \le 1 \Big\}.$$

Moreover, there exists a constant $C_1 > 0$ such that $\rho(\xi) \le C_1 |\xi|$ and $\rho^0(\sigma) \ge \frac{1}{C_1} |\sigma|$ for any ξ, σ. In particular $\rho^0(\sigma) = 0$ if and only if $\sigma = 0$.

Proof. Function $\bar{j}$ is a point-wise supremum of a family of convex non-negative functions: $\{ j(H, \cdot\,) : H \in \mathcal{K}_1 \}$ and thus it is convex and non-negative itself. The value $\bar{j}(\xi)$ is finite since for every ξ the function $j(\,\cdot\,, \xi)$ is real valued, non-negative and concave. Together with convexity and $\dim(V) < \infty$ this fact yields continuity of $\bar{j}$. The p-homogeneity property is straightforward to check and then the formulas (3.9), (3.10) are a standard

result, cf. [140, Corollary 15.3.1]. Since ρ is a real valued gauge the origin is an element of interior of the sublevel set $\{\xi \in V : \rho(\xi) \leq 1\}$ therefore this set contains Euclidean ball $B(0, 1/C_1)$ for C_1 big enough – the gauge generated by this ball is $C_1|\xi|$ and the estimates in the last statement of the proposition follow. □

From the definition of polar function follows that for any $\xi, \sigma \in V$ we have the inequality:

$$\langle \xi, \sigma \rangle \leq \rho(\xi)\, \rho^0(\sigma). \tag{3.11}$$

We say that the pair ξ, σ satisfies the *extremality condition* whenever above there is an equality. The next short result shows a simple link between extremality condition and the subdifferentiability relation $\sigma \in \partial\bar{j}(\xi)$:

Proposition 3.12. *For any pair $\xi, \sigma \in V$ we have the implication:*

$$\sigma \in \partial\bar{j}(\xi) \qquad \Rightarrow \qquad \langle \xi, \sigma \rangle = \rho(\xi)\, \rho^0(\sigma).$$

The equivalence

$$\langle \xi, \sigma \rangle = \rho(\xi)\, \rho^0(\sigma) \qquad \Leftrightarrow \qquad \sigma \in \partial\bar{j}(\xi)$$

holds in the class of ξ, σ satisfying $\big(\rho(\xi)\big)^p = \big(\rho^0(\sigma)\big)^{p'}$ and only in this class.

Proof. By (3.11) and the Young's inequality we obtain

$$\langle \xi, \sigma \rangle \leq \rho(\xi)\, \rho^0(\sigma) \leq \frac{1}{p}\big(\rho(\xi)\big)^p + \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = \bar{j}(\xi) + \bar{j}^*(\sigma).$$

The first implication is due to (3.5). The second statement follows from the fact that the Young's inequality above is an equality if and only if $\big(\rho(\xi)\big)^p = \big(\rho^0(\sigma)\big)^{p'}$. □

Although functions $\bar{j}$, ρ are real valued, the functions $\bar{j}^*, \rho^0$, in general, may admit infinite values. In order to rule this out we require an additional assumption:

- ($\mathrm{A}_{\mathrm{ellip}}$) For every non-zero $\xi \in V$ there exists $H \in \mathcal{K}$ such that $j(H, \xi) > 0$.

Indeed, if ($\mathrm{A}_{\mathrm{ellip}}$) does not hold then there exists $\xi_0 \neq 0$ such that $\bar{j}(\xi_0) = 0$ and thus for any $\sigma_0 \in V$ satisfying $\langle \xi_0, \sigma_0 \rangle > 0$ we have $\bar{j}^*(\sigma) \geq \sup_{t \geq 0}\{\langle t\,\xi_0, \sigma_0 \rangle - \bar{j}(t\,\xi_0)\} = \sup_{t \geq 0} t\langle \xi_0, \sigma_0 \rangle = \infty$. On the other hand we see that

$$(\mathrm{A}_{\mathrm{ellip}}) \qquad \Leftrightarrow \qquad \bar{j}(\xi) = 0 \ \text{ if and only if } \ \xi = 0, \tag{3.12}$$

therefore ($\mathrm{A}_{\mathrm{ellip}}$) shall be called *the ellipticity condition*. Condition ($\mathrm{A}_{\mathrm{ellip}}$) will *not* be assumed in the work unless clearly declared. Naturally, (3.12) holds with $\bar{j}$ replaced by ρ hence we see that, up to the symmetry condition, ($\mathrm{A}_{\mathrm{ellip}}$) implies that ρ satisfies the properties of a norm and it is straightforward that:

Proposition 3.13. *The following conditions are equivalent:*

(i) the ellipticity assumption ($\mathrm{A}_{\mathrm{ellip}}$) *is satisfied;*

(ii) the mutually polar functions ρ, ρ^0 are real valued, continuous, convex positively one-homogeneous functions that for some positive constants C_1, C_2 satisfy

$$C_2|\xi| \le \rho(\xi) \le C_1|\xi| \qquad \forall\, \xi \in V,$$
$$\frac{1}{C_1}|\sigma| \le \rho^0(\sigma) \le \frac{1}{C_2}|\sigma| \qquad \forall\, \sigma \in V^*$$

(ρ, ρ^0 are mutually dual norms on V, $V^* \equiv V$ if and only if they are symmetric).

The two mutually polar gauges ρ, ρ^0 are of fundamental importance as they will encode the anisotropy design in the pair of convex problems $(\mathcal{P}), (\mathcal{P}^*)$, see Chapter 4. We recall that $\bar{j} = \frac{1}{p}(\rho(\,\cdot\,))^p$ where $\bar{j}$ is defined via a maximization problem. On the other hand we have $\bar{j}^* = \frac{1}{p'}(\rho^0(\,\cdot\,))^{p'}$ where $\bar{j}^* = \big(\bar{j}\big)^*$. It is natural to ask how does the function $\bar{j}^*$ depends on the family of functions $j^*(H, \,\cdot\,)$ and whether this dependence is again based on a programming problem with respect to H. Since $j^*(\,\cdot\,, \sigma)$ is convex for each σ we look into the minimization problem:

$$(\underline{j}^*)(\sigma) := \inf_{H \in \mathcal{K}_1} j^*(H, \sigma), \tag{3.13}$$

where the symbol $(\underline{j}^*)$ ought to represent minimum value of j^* with respect to H. Since $j^*(\,\cdot\,, \sigma)$ is convex and lower semi-continuous for any σ and $\mathcal{K}_1$ is compact the minimum in (3.13) is achieved whenever $(\underline{j}^*)(\sigma) < \infty$, i.e. whenever $j^*(\,\cdot\,, \sigma)$ is proper. As we shall see this it is not the case without the additional assumption ($\mathrm{A}_{\mathrm{ellip}}$). Nevertheless, if $(\underline{j}^*)(\sigma) = \infty$ then $j^*(H, \sigma) = \infty = (\underline{j}^*)(\sigma)$ for any $H \in \mathcal{K}_1$. Therefore the set

$$\underline{\mathcal{K}}_1(\sigma) := \Big\{ H \in \mathcal{K}_1 : j^*(H, \sigma) = (\underline{j}^*)(\sigma) \Big\}$$

is a non-empty compact convex subset of $\mathcal{K}_1$ for any $\sigma \in V$.

At this point properties of $(\underline{j}^*)(\sigma)$ are not clear. The next result establishes the several relations between the functions and sets introduced in this section, including the equality $\bar{j}^* = (\,\bar{j}\,)^* = (\underline{j}^*)$. We will also further investigate the equality $\langle \xi, \sigma \rangle = \rho(\xi)\, \rho^0(\sigma)$ to give a link that is deeper than Proposition 3.12: we will show that the extremality equality is connected to the constitutive law $\sigma \in \partial j(\breve{H}, \xi)$ and optimality of anisotropy: $\breve{H} \in \underline{\mathcal{K}}_1(\sigma)$.

Theorem 3.14. *The following statements hold:*

(i) For every $\sigma \in V$

$$(\underline{j}^*)(\sigma) = \inf_{H \in \mathcal{K}_1} j^*(H, \sigma) \;\; = \;\; \bar{j}^*(\sigma) = \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'}. \tag{3.14}$$

In particular $(\underline{j}^*) : V \to \overline{\mathbb{R}}$ is convex, proper and lower semi-continuous.

(ii) for a triple $(\xi, \sigma, \breve{H}) \in V \times V \times \mathcal{K}$ satisfying: $\rho(\xi) \le 1$, $\sigma \ne 0$ and $\rho^0(\sigma) < \infty$, $\breve{H} \in \mathcal{K}_1$ the following conditions are equivalent:

(1) there hold the extremality conditions:

$$\langle \xi, \sigma \rangle = \rho^0(\sigma) \qquad \text{and} \qquad \breve{H} \in \underline{\mathcal{K}}_1(\sigma);$$

(2) the constitutive law is satisfied:

$$\frac{1}{\rho^0(\sigma)}\, \sigma \in \partial j\big(\breve{H}, \xi\big). \tag{3.15}$$

Moreover, for each of the conditions (1), (2) to be true it is necessary that $\rho(\xi) = 1$.

(iii) The following implication is true for every $\xi \in V$ *and non-zero* $\sigma \in V$

$$\langle \xi, \sigma \rangle = \rho(\xi)\, \rho^0(\sigma) \qquad \Rightarrow \qquad \underline{\mathcal{K}}_1(\sigma) \subset \bar{\mathcal{K}}_1(\xi).$$

Proof. For the proof of statement (i) we fix a non-zero element $\sigma \in V$, then directly by definition of the convex conjugate (3.4) we obtain a min-max problem:

$$\begin{aligned}(\underline{j}^*)(\sigma) &= \inf_{H \in \mathcal{K}_1} j^*(H, \sigma) = \inf_{H \in \mathcal{K}_1} \sup_{\zeta \in V} \Big\{ \langle \zeta, \sigma \rangle - j(H, \zeta) \Big\} = \sup_{\zeta \in V} \min_{H \in \mathcal{K}_1} \Big\{ \langle \zeta, \sigma \rangle - j(H, \zeta) \Big\} \\ &= \sup_{\zeta \in V} \Big\{ \langle \zeta, \sigma \rangle - \max_{H \in \mathcal{K}_1} j(H, \zeta) \Big\} = \sup_{\zeta \in V} \Big\{ \langle \zeta, \sigma \rangle - \bar{j}(\zeta) \Big\} = \bar{j}^*(\sigma). \end{aligned} \tag{3.16}$$

Above, to interchange the order of inf and sup we have used Ky Fan's theorem (cf. the first part of Theorem 2.5). Statement (i) is thus proved.

We move on to prove statement (ii). We fix $\xi \in V$ with $\rho(\xi) \leq 1$, non-zero $\sigma \in V$ with $\rho^0(\sigma) < \infty$ and $\breve{H} \in \mathcal{K}_1$. Since $\rho^0(\sigma) > 0$ (cf. the last statement in Proposition 3.11) it is not restrictive to assume that $\rho^0(\sigma) = 1$. Let us first assume that (1) holds, i.e. that $\langle \xi, \sigma \rangle = \rho^0(\sigma) = 1$ and $\breve{H}$ is an element of the non-empty set $\underline{\mathcal{K}}_1(\sigma)$, so that $j^*(\breve{H}, \sigma) = \min_{H \in \mathcal{K}_1} j^*(H, \sigma) = \bar{j}^*(\sigma) < \infty$.

Since $\langle \xi, \sigma \rangle = \rho^0(\sigma) > 0$ and $\langle \xi, \sigma \rangle \leq \rho(\xi)\rho^0(\sigma)$ at the same time we must have $\rho(\xi) = 1$. As a result $\langle \xi, \sigma \rangle = \rho(\xi)\rho^0(\sigma)$ while $\rho(\xi) = \rho^0(\sigma)$ therefore $\sigma \in \partial \bar{j}(\xi)$ by Proposition 3.12. As a result we see that $\zeta = \xi$ solves the last maximization problem in (3.16) and consequently it solves all three last problems in ζ. On the other hand, condition $\breve{H} \in \underline{\mathcal{K}}_1(\sigma)$ means that $j^*(\breve{H}, \sigma) = \min_{H \in \mathcal{K}_1} j^*(H, \sigma) = \bar{j}^*(\sigma)$, i.e. $\breve{H}$ solves the first two minimization problems in (3.16). We thus infer that $(\breve{H}, \xi)$ is a saddle point for the functional $(H, \zeta) \mapsto \langle \zeta, \sigma \rangle - j(H, \zeta)$, i.e.

$$\begin{aligned}\langle \xi, \sigma \rangle - j(\breve{H}, \xi) &= \max_{\zeta \in V} \min_{H \in \mathcal{K}_1} \Big\{ \langle \zeta, \sigma \rangle - j(H, \zeta) \Big\} = \max_{\zeta \in V} \Big\{ \langle \zeta, \sigma \rangle - j(\breve{H}, \zeta) \Big\} = j^*(\breve{H}, \sigma) \\ &= \min_{H \in \mathcal{K}_1} \max_{\zeta \in V} \Big\{ \langle \zeta, \sigma \rangle - j(H, \zeta) \Big\} = \min_{H \in \mathcal{K}_1} \Big\{ \langle \xi, \sigma \rangle - j(H, \xi) \Big\} = \langle \xi, \sigma \rangle - \bar{j}(\xi), \end{aligned}$$

furnishing two equalities: $\langle \xi, \sigma \rangle - j(\breve{H}, \xi) = j^*(\breve{H}, \sigma)$ and $\langle \xi, \sigma \rangle - j(\breve{H}, \xi) = \langle \xi, \sigma \rangle - \bar{j}(\xi)$ from which we infer that $\sigma \in \partial j(\breve{H}, \xi)$ and, respectively, $\breve{H} \in \bar{\mathcal{K}}_1(\xi)$. The former conclusion gives the implication (1) $\Rightarrow$ (2) while the latter, since $\breve{H}$ was arbitrary element of $\underline{\mathcal{K}}_1(\sigma)$ and $\langle \xi, \sigma \rangle = \rho^0(\sigma)$, establishes the implication in point (iii) for the case when $\rho(\xi) = 1$ and $\rho^0(\sigma) < \infty$. For ξ satisfying $\rho(\xi) > 0$ the implication follows by the fact that $\bar{\mathcal{K}}_1(\xi) = \bar{\mathcal{K}}_1(t\,\xi)$ for any $t > 0$ while in the case when $\rho(\xi) = 0$ we have $\underline{\mathcal{K}}_1(\xi) = \mathcal{K}_1$. Finally $\langle \xi, \sigma \rangle < \rho(\xi)\rho^0(\sigma)$ whenever $\rho^0(\sigma) = \infty$ and statement (iii) is established.

For the implication (2) $\Rightarrow$ (1) we assume that for a triple $\xi, \sigma, \breve{H}$ with $\rho(\xi) \leq 1$, $\rho^0(\sigma) = 1$, $\breve{H} \in \mathcal{K}_1$ the constitutive law (3.15) is satisfied. Then, by Proposition 3.3 we have the repartition of energy: $\langle \xi, \sigma \rangle = p\, j(\breve{H}, \xi) = p'\, j^*(\breve{H}, \sigma)$. The following chain of estimates holds:

$$1 = \rho^0(\sigma) = p' \left(\frac{1}{p'} \big(\rho^0(\sigma) \big)^{p'} \right) \leq p' j^*(\breve{H}, \sigma) = \langle \xi, \sigma \rangle = p\, j(\breve{H}, \xi) \leq p \left(\frac{1}{p} \big(\rho(\xi) \big)^{p} \right) \leq 1$$

and therefore all the inequalities above are in fact equalities; in particular we have

$$\langle \xi, \sigma \rangle = \rho^0(\sigma) = \rho(\xi) = 1, \qquad \bar{j}^*(\sigma) = \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = j^*(\breve{H}, \sigma) \quad \Rightarrow \quad \breve{H} \in \underline{\mathcal{K}}_1(\sigma),$$

which proves implication (2) $\Rightarrow$ (1) and the "moreover part" of point (ii). $\square$

Remark 3.15. In general, for a pair of non-zero ξ, σ satisfying the extremality equality $\langle \xi, \sigma \rangle = \rho(\xi)\, \rho^0(\sigma)$ the set $\underline{\mathcal{K}}_1(\sigma)$ may be a proper subset of $\bar{\mathcal{K}}_1(\xi)$, which shall be demonstrated in several examples in Section 3.3, see for instance Example 3.18. The set $\underline{\mathcal{K}}_1(\sigma)$ will appear as the more important one as in the optimal bodies point-wise the anisotropy $\mathscr{H}(x)$ will have to be optimally chosen for the stress (terminology from elasticity) $\sigma(x)$, i.e. $\mathscr{H}(x)$ will have to be an element of $\underline{\mathcal{K}}_1\big(\sigma(x)\big)$.

3.3. Case study: examples of anisotropy design settings

In the previous section we have formulated the problem of anisotropy design the we shall summarize now. It boils down to a pair of finite dimensional programming problems:

$$\frac{1}{p}\big(\rho(\xi)\big)^p = \bar{j}(\xi) = \max_{H \in \mathcal{K}_1} j(H, \xi), \qquad \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = \bar{j}^*(\sigma) = \inf_{H \in \mathcal{K}_1} j^*(H, \sigma).$$

where $\mathcal{K}_1 = \big\{H \in \mathcal{K} \,:\, c(H) \leq 1\big\}$. The problems are constituted by three parameters: $\mathcal{K}, j, c$ and this section is devoted to study of those programs for a number of different settings of $\mathcal{K}, j, c$. Apart from computing the gauges ρ, ρ^0 we will also specify conditions for the extremality equality $\langle \xi, \sigma \rangle = \rho(\xi)\, \rho^0(\sigma)$. Finally, we will characterize the optimal sets

$$\bar{\mathcal{K}}_1(\xi) = \Big\{H \in \mathcal{K}_1 : j(H, \xi) = \bar{j}(\xi)\Big\}, \qquad \underline{\mathcal{K}}_1(\sigma) := \Big\{H \in \mathcal{K}_1 : j^*(H, \sigma) = \bar{j}^*(\sigma)\Big\},$$

where we used the fact that $\bar{j}^* = (\underline{j}^*)$. In most of the examples considered here determining the set $\bar{\mathcal{K}}_1(\xi)$ directly is far easier than in case of $\underline{\mathcal{K}}_1(\sigma)$ since for the latter set we must work with the conjugate function j^* whose explicit form may be cumbersome. For instance, in case when $\mathcal{K} \subset \mathscr{L}_+(V)$ and $j(H, \xi) = \frac{1}{2}\langle H\xi, \xi \rangle$ we have (provided that $j^*(H, \sigma) < \infty$) $j^*(H, \sigma) = \frac{1}{2}\langle H^{-1}\sigma, \sigma \rangle$ where H^{-1} is the *pseudo inverse operator*. In almost all of the examples a different strategy will be undertaken where after computing $\bar{\mathcal{K}}_1(\xi)$ we shall utilize statements (ii), (iii) in Theorem 3.14 to characterize $\underline{\mathcal{K}}_1(\sigma)$ through the extremality equation $\langle \xi, \sigma \rangle = \rho(\xi)\, \rho^0(\sigma)$ and the constitutive law $\sigma / \rho^0(\sigma) \in \partial j\big(H, \xi\big)$.

3.3.1. The linear constitutive law

In the present section the set $\mathcal{K}$ will be a closed convex cone contained in the set $\mathscr{L}_+(V)$ of positive semi-definite operators, while the cost function will be taken as the trace operator. We shall only consider the energy function j that generates the linear constitutive law. More precisely:

$$\mathcal{K} \subset \mathscr{L}_+(V), \qquad j(H, \xi) = \frac{1}{2}\langle H\xi, \xi \rangle \quad (p = 2), \qquad c(H) = \operatorname{Tr} H.$$

The examples shall thus vary only with respect to the set $\mathcal{K}$. As already mentioned in Example 3.4 the assumptions (A1)-(A4) are easy to verify for the function j above, similarly $c = \mathrm{Tr}$ is admissible according to Example 3.9. Validity of the extra ellipticity assumption ($\mathrm{A}_{\mathrm{ellip}}$) virtually depends on the set $\mathcal{K}$ as it must be "large enough". In all of the examples considered in Section 3.3.1 the functions ρ, ρ^0 will turn out to be norms thus validating ($\mathrm{A}_{\mathrm{ellip}}$) by virtue of Proposition 3.13.

The first example will be given in the abstract form in order to fit with both elasticity and conductivity context, i.e. for $V = \mathcal{S}^{d\times d}$ and $V = \mathbb{R}^d$ respectively. The subsequent examples will focus on elasticity only, i.e. we shall write $\mathcal{S}^{d\times d}$ in place of V and V^*.

Example 3.16 (**Anisotropic Material Design problem**)**.** We consider the anisotropy design problem that in the context of the Free Material Design is mostly discussed in the literature, cf. [136], [20], [64]: the *Anisotropic Material Design* (AMD) where we admit arbitrary anisotropy of the material, i.e.

$$\mathcal{K} = \mathscr{L}_+(V). \tag{3.17}$$

For each $H \in \mathcal{K}_1 = \big\{H \in \mathcal{K} \,:\, \mathrm{Tr}\, H \le 1\big\}$ we may write down an estimate

$$j(H,\xi) = \frac{1}{2}\langle H\,\xi,\xi\rangle \le \frac{1}{2}\left(\max_{i\in\{1,\dots,N\}} \lambda_i(H)\right)|\xi|^2 \le \frac{1}{2}\left(\mathrm{Tr}\, H\right)|\xi|^2 \le \frac{1}{2}\,|\xi|^2$$

($|\xi| = \langle\xi,\xi\rangle^{1/2}$ denotes the Euclidean norm of ξ) and therefore $\frac{1}{2}\big(\rho(\xi)\big)^2 = \bar{j}(\xi) \le \frac{1}{2}\,|\xi|^2$. On the other hand, for a fixed non-zero $\xi \in V$ we may define

$$\bar{H}_\xi = \frac{\xi}{|\xi|}\otimes\frac{\xi}{|\xi|} \tag{3.18}$$

that is a rank-one operator with the single non-zero eigenvalue being equal to 1 and the corresponding unit eigenvector $\xi/|\xi|$; obviously we have $\mathrm{Tr}\,\bar{H}_\xi = 1$ and $j(\bar{H}_\xi,\xi) = \frac{1}{2}\big\langle \bar{H}_\xi\,\xi,\xi\big\rangle = \frac{1}{2}\,|\xi|^2$, which shows that in fact $\bar{j}(\xi) = j(\bar{H}_\xi,\xi) = \frac{1}{2}\,|\xi|^2$. As a result

$$\rho(\xi) = |\xi|, \qquad \frac{\xi}{|\xi|}\otimes\frac{\xi}{|\xi|} \in \bar{\mathcal{K}}_1(\xi).$$

It is straightforward to check that for a non-zero ξ the tensor $\bar{H}_\xi$ is the unique element of the set $\bar{\mathcal{K}}_1(\xi)$. Indeed, if $\xi/|\xi|$ is kept as one of the eigenvectors of a chosen $H \in \mathcal{K}_1$ and $\lambda_1(H)$ is the corresponding eigenvalue, then $H \neq \bar{H}_\xi$ means that $\lambda_1(H) < 1$ yielding $j(H,\xi) < \frac{1}{2}|\xi|^2$. One may easily verify that we obtain a similar result whenever $\xi/|\xi|$ is not one of the eigenvectors of H.

It is well known that the polar ρ^0 to the Euclidean norm $\rho = |\cdot|$ is again this very norm. Furthermore it is straightforward that

$$\langle\xi,\sigma\rangle = \rho(\xi)\,\rho^0(\sigma) = |\xi|\,|\sigma| \qquad \Leftrightarrow \qquad t_1\xi = t_2\sigma \ \text{ for some } \ t_1,t_2 \ge 0.$$

Next, for non-zero σ the point (iii) of Theorem 3.14 furnishes:

$$\xi_\sigma = t\,\sigma \ \text{ for } \ t>0 \quad \Rightarrow \quad \underline{\mathcal{K}}_1(\sigma) \subset \bar{\mathcal{K}}_1(\xi_\sigma) = \left\{\frac{\xi_\sigma}{|\xi_\sigma|}\otimes\frac{\xi_\sigma}{|\xi_\sigma|}\right\} = \left\{\frac{\sigma}{|\sigma|}\otimes\frac{\sigma}{|\sigma|}\right\}$$

therefore, since $\underline{\mathcal{K}}_1(\sigma)$ is non-empty and $\bar{\mathcal{K}}_1(\xi_\sigma)$ is a singleton, we have obtained

$$\rho^0(\sigma) = |\sigma|, \qquad \underline{\mathcal{K}}_1(\sigma) = \left\{\frac{\sigma}{|\sigma|}\otimes\frac{\sigma}{|\sigma|}\right\}. \tag{3.19}$$

The latter results were obtained in [64] by solving the problem $\min_{H\in\mathcal{K}_1} j^*(H,\sigma)$ directly. △

As announced, in the sequel of Section 3.3 we fix our focus on the physical context of d-dimensional elasticity where $V=\mathcal{S}^{d\times d}$, $d=2$ or $d=3$. The variety of important sets $\mathcal{K}\subset\mathscr{L}_+(\mathcal{S}^{d\times d})$ is much wider than in the case of conductivity where $\mathcal{K}\subset\mathcal{S}^{d\times d}_+$. The elements $H\in\mathscr{L}_+(\mathcal{S}^{d\times d})$ may be identified with Hooke tensors and, abusing the terminology, thus they shall be called in the examples.

Example 3.17 (**Isotropic Material Design problem**)**.** The following variant of anisotropy design is known as the *Isotropic Material Design* problem (IMD), cf. [62, 68, 67]:

$$\mathcal{K}=\mathcal{K}_{iso}:=\left\{\mathscr{K}\left(\frac{1}{d}\,\mathrm{I}\otimes\mathrm{I}\right)+2\mathscr{G}\left(\mathrm{Id}-\frac{1}{d}\,\mathrm{I}\otimes\mathrm{I}\right)\ :\ \mathscr{K},\mathscr{G}\geq 0\right\},$$

while $j(H,\xi)=\frac{1}{2}\langle H\xi,\xi\rangle$ and $c(H)=\operatorname{Tr}H$. The set $\mathcal{K}_{iso}$ is (relatively) a two-dimensional closed convex cone of isotropic Hooke tensors in a d-dimensional body, $d\in\{2,3\}$. For any $H\in\mathcal{K}_{iso}$ and $\xi\in\mathcal{S}^{d\times d}$ we have

$$j(H,\xi)=\frac{1}{2}\left(\mathscr{K}\,|\operatorname{Tr}\xi|^2+2\mathscr{G}\,|\operatorname{dev}\xi|^2\right)\quad\text{where}\quad\operatorname{dev}\xi:=\xi-\frac{1}{d}\,(\operatorname{Tr}\xi)\,\mathrm{I}=\Big(\mathrm{Id}-\frac{1}{d}\,\mathrm{I}\otimes\mathrm{I}\Big)\,\xi$$

and $|\operatorname{dev}\xi|$ stands for the Euclidean norm of the matrix. It is well established that $H\in\mathcal{K}_{iso}$ has a single eigenvalue $d\mathscr{K}$ and $N(d)-1$ eigenvalues $2\mathscr{G}$ where $N(d)=\dim(\mathcal{S}^{d\times d})=d\,(d+1)/2$. Therefore $\operatorname{Tr}H=d\,\mathscr{K}+\big(N(d)-1\big)\,2\mathscr{G}$ and upon introducing auxiliary variables $A_1=d\,\mathscr{K}$ and $A_2=2\big(N(d)-1\big)\,\mathscr{G}$ we obtain $\operatorname{Tr}H=A_1+A_2$ and

$$j(H,\xi)=\frac{1}{2}\left(A_1\left(\frac{|\operatorname{Tr}\xi|}{\sqrt{d}}\right)^2+A_2\left(\frac{|\operatorname{dev}\xi|}{\sqrt{N(d)-1}}\right)^2\right).$$

As a result

$$\bar{j}(\xi)=\max_{H\in\mathcal{K}_1}j(H,\xi)=\max_{A_1,A_2\geq 0}\big\{j(H,\xi)\ :\ A_1+A_2\leq 1\big\}=\frac{1}{2}\big(\rho(\xi)\big)^2$$

where

$$\rho(\xi)=\max\left\{\frac{|\operatorname{Tr}\xi|}{\sqrt{d}},\frac{|\operatorname{dev}\xi|}{\sqrt{N(d)-1}}\right\},\tag{3.20}$$

while, for a non-zero strain $\xi\in\mathcal{S}^{d\times d}$

$$\bar{\mathcal{K}}_1(\xi)=\left\{H\in\mathcal{K}_{iso}\ :\ d\mathscr{K}+2\big(N(d)-1\big)\,\mathscr{G}=1,\quad\begin{array}{rl}\Big(\frac{|\operatorname{Tr}\xi|}{\sqrt{d}}-\rho(\xi)\Big)\,\mathscr{K}&=0,\\ \Big(\frac{|\operatorname{dev}\xi|}{\sqrt{N(d)-1}}-\rho(\xi)\Big)\mathscr{G}&=0\end{array}\right\}.$$

By acknowledging the fact that $\langle\xi,\sigma\rangle=\frac{1}{d}(\operatorname{Tr}\xi)(\operatorname{Tr}\sigma)+\langle\operatorname{dev}\xi,\operatorname{dev}\sigma\rangle$ we arrive at the polar gauge

$$\rho^0(\sigma)=\frac{1}{\sqrt{d}}\,|\operatorname{Tr}\sigma|+\sqrt{N(d)-1}\,|\operatorname{dev}\sigma|\tag{3.21}$$

and the extremality conditions for non-zero ξ,σ follow:

$$\langle\xi,\sigma\rangle=\rho(\xi)\,\rho^0(\sigma)\qquad\Leftrightarrow\qquad\left\{\begin{array}{l}\operatorname{Tr}\sigma=|\operatorname{Tr}\sigma|\,\frac{\operatorname{Tr}\xi}{\sqrt{d}\,\rho(\xi)},\\ \operatorname{dev}\sigma=|\operatorname{dev}\sigma|\,\frac{\operatorname{dev}\xi}{\sqrt{N(d)-1}\,\rho(\xi)}.\end{array}\right.\tag{3.22}$$

In order to characterize the set of optimal Hooke tensors $\underline{\mathcal{K}}_1(\sigma)$ we use Theorem 3.14. For a fixed stress $\sigma \in \mathcal{S}^{d\times d}$ let $\xi = \xi_\sigma \in \mathcal{S}^{d\times d}$ be a strain tensor such that $\rho(\xi_\sigma) = 1$ and the extremality conditions (3.22) hold. Then, according to point (ii) of Theorem 3.14 a Hooke tensor $H \in \mathcal{K}_1$ belongs to $\underline{\mathcal{K}}_1(\sigma)$ if and only if the constitutive law $\sigma/\rho^0(\sigma) = \partial j(H, \xi_\sigma)$ holds. For our j the law is linear: $\sigma/\rho^0(\sigma) = H\,\xi_\sigma$ and, considering that $\sigma = \frac{1}{d}\,(\operatorname{Tr}\sigma)\,\mathrm{I} + \operatorname{dev}\sigma$ and acknowledging (3.22), the optimality criterion reads

$$\frac{1}{\rho^0(\sigma)}\left(\frac{1}{d}\,|\operatorname{Tr}\sigma|\,\frac{\operatorname{Tr}\xi_\sigma}{\sqrt{d}}\,\mathrm{I} + |\operatorname{dev}\sigma|\,\frac{\operatorname{dev}\xi_\sigma}{\sqrt{N(d)-1}}\right) = \mathscr{K}\,(\operatorname{Tr}\xi_\sigma)\,\mathrm{I} + 2\mathscr{G}\operatorname{dev}\xi_\sigma.$$

It is easy to see that for any σ the tensor ξ_σ may be chosen so that both $\operatorname{Tr}\xi_\sigma \neq 0$ and $\operatorname{dev}\xi_\sigma \neq 0$ and then comparing the left and right-hand side above yields

$$\underline{\mathcal{K}}_1(\sigma) = \left\{H \in \mathcal{K}_{iso} \;:\; \mathscr{K} = \frac{1}{d\sqrt{d}}\,\frac{|\operatorname{Tr}\sigma|}{\rho^0(\sigma)},\; \mathscr{G} = \frac{1}{2\sqrt{N(d)-1}}\,\frac{|\operatorname{dev}\sigma|}{\rho^0(\sigma)}\right\}. \tag{3.23}$$

We notice that $\underline{\mathcal{K}}_1(\sigma)$ is always a singleton for non-zero σ, while $\bar{\mathcal{K}}_1(\xi)$ may be a one dimensional affine subset of $\mathcal{K}_{iso}$, provided $|\operatorname{Tr}\xi|/\sqrt{d} = |\operatorname{dev}\xi|/\sqrt{N(d)-1} = \rho(\xi) \neq 0$. $\triangle$

The next anisotropy setting was put forward by the present author to simulate design of a fibrous-like material:

Example 3.18 (**Fibrous Material Design problem**)**.** We define a set

$$\mathcal{K}_{axial} := \Big\{H = a\,(\tau\otimes\tau)\otimes(\tau\otimes\tau) \;:\; a \geq 0,\; \tau \in S^{d-1}\Big\}.$$

where $S^{d-1} = \big\{\tau \in \mathbb{R}^d : |\tau| = 1\big\}$ denotes the unit sphere in $\mathbb{R}^d$. Above $\tau\otimes\tau$ denotes a rank-one matrix in $\mathcal{S}^{d\times d}$ while $(\tau\otimes\tau)\otimes(\tau\otimes\tau)$ is an operator in $\mathscr{L}_+(\mathcal{S}^{d\times d})$ being a tensor product of elements in $\mathcal{S}^{d\times d}$. As mentioned in Example 3.1 the set $\mathscr{L}_+(\mathcal{S}^{d\times d})$ is isomorphic to a set of 4th order Hooke tensor, hence it is justified to write $\tau\otimes\tau\otimes\tau\otimes\tau$ instead, which we shall do henceforward.

The elements of $\mathcal{K}_{axial}$ ought to represent uni-axial Hooke tensors that correspond to material made from a family of fibres aligned along direction τ. However, the set $\mathcal{K}_{axial}$ is a closed cone that is not convex. We thus propose the setting of the *Fibrous Material Design* (FibMD) where we choose

$$\mathcal{K} = \mathcal{K}_{fib} := \operatorname{co}\big(\mathcal{K}_{axial}\big)$$

where co is the convex hull; we still keep $j(H,\xi) = \frac{1}{2}\langle H\xi, \xi\rangle$ and $c(H) = \operatorname{Tr} H$. The set $\mathcal{K}_{fib}$ is designed to simulate a fibrous material that is a composite of uni-axial materials. Clearly the set $\mathcal{K}$ is a convex cone; it may be easily checked that it is also closed as convex hull of any closed cone in a finite dimensional space is closed.

We first observe that for each $H \in \mathcal{K}_{axial}$ with $c(H) \leq 1$, i.e. with $\operatorname{Tr} H = a \leq 1$, there holds

$$j(H,\xi) = \frac{1}{2}\langle H\,\xi, \xi\rangle = \frac{a}{2}\,\big(\langle \xi, \tau\otimes\tau\rangle\big)^2 \leq \frac{1}{2}\,\left(\max_{i\in\{1,\ldots,d\}} |\lambda_i(\xi)|\right)^2 \tag{3.24}$$

and at the same time

$$j(\bar{H}_\xi, \xi) = \frac{1}{2}\,\left(\max_{i\in\{1,\ldots,d\}} |\lambda_i(\xi)|\right)^2 \quad \text{for} \quad \bar{H}_\xi = \bar{v}(\xi)\otimes\bar{v}(\xi)\otimes\bar{v}(\xi)\otimes\bar{v}(\xi) \tag{3.25}$$

where $\bar{v}(\xi)$ is any unit eigenvector of ξ corresponding to an eigenvalue of maximal absolute value.

Let us now take any $\tilde{H} \in \mathrm{co}\big(\mathcal{K}_{axial}\big)$, namely (thanks to the cone structure) $\tilde{H} = \sum_{i=1}^m \alpha_i H_i$ for some $\alpha_i \geq 0$ and $H_i \in \mathcal{K}_{axial}$ with $c(H_i) > 0$. Since both $c = \mathrm{Tr}$ and $j(\,\cdot\,,\xi)$ are linear there holds $c(\tilde{H}) = \sum_{i=1}^m \alpha_i\, c\,(H_i)$ and

$$\begin{aligned} j(\tilde{H},\xi) &= \sum_{i=1}^m \alpha_i\, j(H_i,\xi) = \sum_{i=1}^m \alpha_i\, c(H_i)\; j\left(\frac{H_i}{c(H_i)},\xi\right) \\ &\leq \left(\sup_{\substack{H\in\mathcal{K}_{axial}\\ c(H)\leq 1}} j(H,\xi)\right) \sum_{i=1}^m \alpha_i\, c(H_i) = \left(\sup_{\substack{H\in\mathcal{K}_{axial}\\ c(H)\leq 1}} j(H,\xi)\right)\, c(\tilde{H}). \end{aligned}$$

By recalling (3.24) and (3.25) we arrive at

$$\bar{j}(\xi) = \max_{\substack{H\in\mathcal{K}_{fib}\\ c(H)\leq 1}} j(H,\xi) = \max_{\substack{H\in\mathcal{K}_{axial}\\ c(H)\leq 1}} j(H,\xi) = \frac{1}{2}\,\left(\max_{i\in\{1,\dots,d\}} |\lambda_i(\xi)|\right)^2, \tag{3.26}$$

where the first equality is by definition of $\bar{j}$; moreover

$$\bar{\mathcal{K}}_1(\xi) = \mathrm{co}\left(\left\{\bar{v}\otimes\bar{v}\otimes\bar{v}\otimes\bar{v}\;:\; \bar{v} \text{ is any eigenvector } v_i(\xi) \text{ with maximal } |\lambda_i(\xi)|\right\}\right). \tag{3.27}$$

As a consequence ρ becomes the spectral norm on the space of symmetric matrices $\mathcal{S}^{d\times d}$; we display it next to the well-established formula for its polar:

$$\rho(\xi) = \max_{i\in\{1,\dots,d\}} |\lambda_i(\xi)|, \qquad \rho^0(\sigma) = \sum_{i=1}^d |\lambda_i(\sigma)|. \tag{3.28}$$

When $\xi \neq 0$ the extremality condition for the pair ρ, ρ^0 may be characterized as follows

$$\langle \xi,\sigma\rangle = \rho(\xi)\,\rho^0(\sigma) \quad \Leftrightarrow \quad \begin{cases} \text{every eigenvector of } \sigma \text{ is an eigenvector of } \xi \text{ and} \\ \lambda_i(\sigma) = \frac{\lambda_i(\xi)}{\rho(\xi)}\,|\lambda_i(\sigma)| \qquad \forall\, i. \end{cases} \tag{3.29}$$

It is thus only left to characterize the set $\underline{\mathcal{K}}_1(\sigma)$; we see that this time we are forced to search the set $\mathcal{K} = \mathcal{K}_{fib}$, instead of just $\mathcal{K}_{axial}$, which turned out to be the case while maximizing $j(\,\cdot\,,\xi)$: indeed, any σ of rank higher than one yields $j^*(H,\sigma) = \infty$ for each $H \in \mathcal{K}_{axial}$. Our strategy for computing $\underline{\mathcal{K}}_1(\sigma)$ will be similar as in Example 3.17.

For a fixed stress $\sigma \in \mathcal{S}^{d\times d}$ let $\xi = \xi_\sigma \in \mathcal{S}^{d\times d}$ be a strain tensor such that $\rho(\xi_\sigma) = 1$ and extremality conditions (3.29) hold. Then, according to point (ii) of Theorem 3.14 a Hooke tensor $H \in \mathcal{K}_1$ belongs to $\underline{\mathcal{K}}_1(\sigma)$ if and only if the constitutive law $\sigma/\rho^0(\sigma) = \partial j(H,\xi_\sigma)$ holds, which here is a linear equation:

$$\frac{\sigma}{\rho^0(\sigma)} = H\,\xi_\sigma. \tag{3.30}$$

With $v_i(\sigma)$ denoting unit eigenvectors of σ for a non-zero stress σ we propose the Hooke tensor

$$\underline{H}_\sigma = \sum_{i=1}^d \frac{|\lambda_i(\sigma)|}{\rho^0(\sigma)}\; v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma) \tag{3.31}$$

that is an element of $\mathcal{K}_1$, i.e. $\underline{H}_\sigma \in \mathcal{K}_{fib}$ and $\operatorname{Tr} \underline{H}_\sigma = 1$. Since the pair ξ_σ, σ satisfies (3.29) each $v_i(\sigma)$ is an eigenvector for ξ_σ and moreover $\langle \xi_\sigma, v_i(\sigma) \otimes v_i(\sigma) \rangle = \operatorname{sign}\big(\lambda_i(\sigma)\big)$, therefore

$$\underline{H}_\sigma \, \xi_\sigma = \sum_{i=1}^{d} \frac{|\lambda_i(\sigma)|}{\rho^0(\sigma)} \operatorname{sign}\big(\lambda_i(\sigma)\big) \; v_i(\sigma) \otimes v_i(\sigma) = \frac{\sigma}{\rho^0(\sigma)},$$

which proves that $\underline{H}_\sigma \in \underline{\mathcal{K}}_1(\sigma)$. The full characterization of the set $\underline{\mathcal{K}}_1(\sigma)$ is difficult to write down for arbitrary d hence further we shall proceed in dimension $d = 2$, where three cases must be examined:

• Case a): the determinant of σ is negative.

In this case σ has two non-zero eigenvalues of opposite sign, let us say: $\lambda_1(\sigma) < 0$ and $\lambda_2(\sigma) > 0$. Therefore the tensor $\xi = \xi_\sigma$ that satisfies $\rho(\xi_\sigma) \leq 1$ and is in the extremal relation (3.29) is uniquely determined: there must hold $\xi_\sigma = -v_1(\sigma) \otimes v_1(\sigma) + v_2(\sigma) \otimes v_2(\sigma)$ where $v_1(\sigma), v_2(\sigma)$ are the respective eigenvectors of σ. According to point (iii) of Theorem 3.14 there must hold $\underline{\mathcal{K}}_1(\sigma) \subset \bar{\mathcal{K}}_1(\xi_\sigma)$ and thus from (3.27) we deduce that each $H \in \bar{\mathcal{K}}_1(\sigma)$ satisfies $H = \sum_{i=1}^{2} \alpha_i \, v_i(\sigma) \otimes v_i(\sigma) \otimes v_i(\sigma) \otimes v_i(\sigma)$ for $\alpha_1 + \alpha_2 = 1$. Then the constitutive law (3.30) enforces $\sigma/\rho^0(\sigma) = -\alpha_1 \, v_1(\sigma) \otimes v_1(\sigma) + \alpha_2 \, v_2(\sigma) \otimes v_2(\sigma)$ and we immediately obtain that $\alpha_i = |\lambda_i(\sigma)|/\rho^0(\sigma)$ and therefore H must coincide with $\underline{H}_\sigma$ from (3.31). In summary, in the case when $d = 2$ and $\det \sigma < 0$ the set $\underline{\mathcal{K}}_1(\sigma)$ is a singleton:

$$\underline{\mathcal{K}}_1(\sigma) = \left\{ \sum_{i=1}^{2} \frac{|\lambda_i(\sigma)|}{\rho^0(\sigma)} \; v_i(\sigma) \otimes v_i(\sigma) \otimes v_i(\sigma) \otimes v_i(\sigma) \right\}, \tag{3.32}$$

while $\bar{\mathcal{K}}_1(\xi_\sigma)$ is the convex hull of $\big\{ v_i(\sigma) \otimes v_i(\sigma) \otimes v_i(\sigma) \otimes v_i(\sigma) \; : \; i \in \{1, 2\} \big\}$.

• Case b): the determinant of σ is positive.

Without loss of generality we may assume that $\lambda_1(\sigma), \lambda_2(\sigma) > 0$. Once again there is only one strain tensor ξ_σ with $\rho(\xi_\sigma) = 1$ and satisfying (3.29): necessarily $\xi_\sigma = \mathrm{I}$. Therefore any vector $\tau \in S^{d-1}$ is an eigenvector of ξ_σ (but not necessarily of σ) with eigenvalue equal to one and thus $\bar{\mathcal{K}}_1(\xi_\sigma) = \operatorname{co}\big\{ \tau \otimes \tau \otimes \tau \otimes \tau \; : \; \tau \in S^{d-1} \big\}$. Therefore the inclusion $\underline{\mathcal{K}}_1(\sigma) \subset \bar{\mathcal{K}}_1(\xi_\sigma)$ merely indicates that for $H \in \underline{\mathcal{K}}_1(\sigma)$ there must hold $H = \sum_{i=1}^{m} \alpha_i \, \tau_i \otimes \tau_i \otimes \tau_i \otimes \tau_i$ where $m \in \mathbb{N}_+$, $\tau_i \in S^{d-1}$, $\alpha_i \geq 0$ and $\sum_{i=1}^{m} \alpha_i = 1$. By plugging this form of H into (3.30) we obtain the characterization for a positive definite σ (recall that $\langle \tau_i \otimes \tau_i, \xi_\sigma \rangle = 1 \;\; \forall \, i$)

$$\underline{\mathcal{K}}_1(\sigma) = \left\{ \sum_{i=1}^{m} \alpha_i \, \tau_i \otimes \tau_i \otimes \tau_i \otimes \tau_i \; : \; \tau_i \in S^{d-1}, \; \sum_{i=1}^{m} \alpha_i = 1, \; \frac{\sigma}{\operatorname{Tr} \sigma} = \sum_{i=1}^{m} \alpha_i \, \tau_i \otimes \tau_i \right\}, \tag{3.33}$$

where we used the fact that $\rho^0(\sigma) = \operatorname{Tr} \sigma$ for any positive semi-definite σ. This set is not a singleton, cf. Remark 3.29 concerning a similar Example 3.28.

• Case c): σ is of rank one.

It is not restrictive to assume that $\lambda_1(\sigma) = 0$, $\lambda_2(\sigma) > 0$ and so $\sigma = \lambda_2(\sigma) \, v_2(\sigma) \otimes v_2(\sigma)$. In this case there are infinitely many ξ_σ such that $\rho(\xi_\sigma) = 1$ and (3.29) holds. We can, however, test (3.30) with only one: $\xi_\sigma := v_2(\sigma) \otimes v_2(\sigma)$ for which $\bar{\mathcal{K}}_1(\xi_\sigma) = \big\{ v_2(\sigma) \otimes v_2(\sigma) \otimes v_2(\sigma) \otimes v_2(\sigma) \big\}$ which is necessarily equal to $\underline{\mathcal{K}}_1(\sigma)$ due to point (iii) of Theorem 3.14.

Eventually, for a rank-one stress σ the set of optimal Hooke tensors may be written as a singleton

$$\underset{\sim}{\mathcal{K}}_1(\sigma) = \left\{ \frac{\sigma}{|\sigma|} \otimes \frac{\sigma}{|\sigma|} \right\}, \tag{3.34}$$

where we used the fact that $\rho^0(\sigma) = |\sigma|$ for σ of rank one. By comparing to Example 3.16 we learn that the AMD and FibMD problems furnish the same optimal Hooke tensor at points where σ is rank-one. △

3.3.2. Non-linear constitutive laws

Having covered a series of examples for $j(H,\xi) = \frac{1}{2}\langle H\xi, \xi\rangle$ we now turn to propose and examine some important functions j that do not generate linear constitutive laws. What is more, we shall find an example of j whose dependence on H is not linear so that the fairly general assumption (A3) is made worthwhile. We shall stick to the elasticity setting only, i.e. $V = \mathcal{S}^{d\times d}$ and $\mathcal{K} \subset \mathscr{L}_+(\mathcal{S}^{d\times d})$.

We start with a simple example that generalizes the function $j(H,\xi) = \frac{1}{2}\langle H\xi, \xi\rangle$ to homogeneity exponents p other than two:

Example 3.19 (**Isotropic Material Design in the case of the power-law**)**.** For $p \in (1,\infty)$ different from 2 one may propose a generalization of the constitutive law of linear elasticity. The conditions (H1)-(H5) can be easily satisfied if one assumes admissible Hooke tensors to be isotropic, i.e. again $H \in \mathcal{K} = \mathcal{K}_{iso}$. For instance we may choose

$$j(H,\xi) = \frac{1}{p}\left(\mathscr{K} |\operatorname{Tr}\xi|^p + 2\mathscr{G}\, |\operatorname{dev}\xi|^p \right)$$

where once more the moduli $\mathscr{K}, \mathscr{G} \geq 0$ identify an isotropic Hooke tensor $H \in \mathcal{K}_{iso}$ by means of (3.2). A similar potential is proposed in [52] and referred to as the *power-law* potential. The authors therein, however, allow to choose different exponents for the two tensor invariants $|\operatorname{Tr}\xi|$ and $|\operatorname{dev}\xi|$, whilst here the assumption (A2) obligates us to apply a common exponent p. The results from Example 3.17 hold here with only slight modifications, for instance

$$\rho(\xi) = \max\left\{ \frac{|\operatorname{Tr}\xi|}{d^{1/p}}, \frac{|\operatorname{dev}\xi|}{\big(N(d)-1\big)^{1/p}} \right\}, \quad \rho^0(\sigma) = \frac{|\operatorname{Tr}\sigma|}{d^{1/p'}} + \big(N(d)-1\big)^{1/p} |\operatorname{dev}\sigma|.$$

△

In elasticity we encounter materials that are sensitive to the sign of the stress: e.g. the material's rigidity is different when in tension and when in compression. A typical example of such a material is concrete. For a non-zero stress tensor $\sigma \in \mathcal{S}^{d\times d}$ we say that material is in pure tension when $\lambda_i(\sigma) \geq 0 \ \ \forall\, i$ and in pure compression when $\lambda_i(\sigma) \leq 0$ $\forall\, i$; the material is in the tension-compression state when neither of the two conditions holds (for $d = 2$ this is equivalent to $\det\sigma < 0$). An extreme case in this class would be a material that cannot withstand tension (or compression) at all: in [90] the authors model a isotropic masonry material by enforcing that σ is negative semi-definite.

Below we propose a consistent way of constructing elastic energy functions for materials that cannot resist compression (or tension). For a chosen cone $\mathcal{K} \subset \mathscr{L}_+(\mathcal{S}^{d\times d})$ we start with any function $j : \mathcal{K} \times \mathcal{S}^{d\times d} \to \mathbb{R}$ that satisfies conditions (A1)-(A4). Then we set:

$$j_+(H,\,\cdot\,) := \big(j^*(H,\,\cdot\,) + \chi_{\mathcal{S}^{d\times d}_+}(\,\cdot\,)\big)^*, \qquad j_-(H,\,\cdot\,) := \big(j^*(H,\,\cdot\,) + \chi_{\mathcal{S}^{d\times d}_-}(\,\cdot\,)\big)^* \tag{3.35}$$

where for a set A in a linear space X symbol χ_A is the indicator function, i.e. $\chi_A(x) = 0$ for $x \in A$ and $\chi_A(x) = \infty$ for $x \in X\backslash A$; the sets $\mathcal{S}^{d\times d}_+$ and $\mathcal{S}^{d\times d}_-$ are the closed convex cones of positive and negative semi-definite symmetric matrices. The functions $j^*(H,\,\cdot\,) + \chi_{\mathcal{S}^{d\times d}_+}$ and $j^*(H,\,\cdot\,) + \chi_{\mathcal{S}^{d\times d}_-}$ are proper convex l.s.c therefore

$$j^*_+(H,\sigma) = j^*(H,\sigma) + \chi_{\mathcal{S}^{d\times d}_+}(\sigma), \qquad j^*_-(H,\sigma) = j^*(H,\sigma) + \chi_{\mathcal{S}^{d\times d}_-}(\sigma).$$

We can already see that the Hooke's law $\sigma \in \partial j_+(H,\xi)$ rules out compression, i.e. stress tensors σ that are not positive semi-definite: indeed, by (3.5) the law is equivalent to equality $\langle \xi, \sigma\rangle = j_+(H,\xi) + j^*_+(H,\sigma)$. Analogously we deduce that $\sigma \in \partial j_-(H,\xi)$ bans tension. We now study the properties of j_+, j_- more deeply:

Proposition 3.20. *Let $j : \mathcal{K} \times \mathcal{S}^{d\times d} \to \mathbb{R}$ be any function that satisfies conditions* (A1)-(A4)*. Then, the function $j_+ : \mathcal{K} \times \mathcal{S}^{d\times d} \to \mathbb{R}$ satisfies* (A1)-(A4) *as well, whilst*

$$j_+(H,\xi) = \inf_{\zeta \in \mathcal{S}^{d\times d}_+} j(H,\xi+\zeta). \tag{3.36}$$

Moreover, $\hat\zeta \in \mathcal{S}^{d\times d}_+$ is a solution of the minimization problem for a given ξ above if and only if there exists $\hat\sigma \in \mathcal{S}^{d\times d}_+$ such that

$$\hat\sigma \in \partial j(H,\xi+\hat\zeta) \qquad \textit{and} \qquad \langle \hat\zeta, \hat\sigma\rangle = 0,$$

when in addition $\hat\sigma \in \partial j_+(H,\xi)$ holds.

The analogous statement holds for j_- provided that above $\mathcal{S}^{d\times d}_+$ is replaced with $\mathcal{S}^{d\times d}_-$.

Proof. We begin by noting that for every non-zero $H \in \mathcal{K}$ we obtain coerciveness of $j^*(H,\,\cdot\,)$ by combining (3.7) and Proposition 3.7:

$$j^*(H,\sigma) = \frac{1}{(c(H))^{1/(p-1)}}\, j^*(\tilde H,\sigma) \geq \frac{1}{(c(H))^{1/(p-1)}}\, \frac{1}{p'} \big(\rho^0(\sigma)\big)^{p'} \geq \frac{|\sigma|^{p'}}{p'\,(C_1)^{p'}\,(c(H))^{1/(p-1)}},$$

where $\tilde H = H/c(H) \in \mathcal{K}_1$.

Next, for a fixed $\xi \in \mathcal{S}^{d\times d}$ we obtain by introducing a Lagrange multiplier ζ:

$$\begin{aligned} j_+(H,\xi) &= \max_{\sigma\in\mathcal{S}^{d\times d}} \Big\{\langle \xi,\sigma\rangle - j^*(H,\sigma) - \chi_{\mathcal{S}^{d\times d}_+}(\sigma)\Big\} = \max_{\sigma\in\mathcal{S}^{d\times d}_+} \Big\{\langle \xi,\sigma\rangle - j^*(H,\sigma)\Big\} \\ &= \max_{\sigma\in\mathcal{S}^{d\times d}} \inf_{\zeta\in\mathcal{S}^{d\times d}_+} \Big\{\langle \xi+\zeta,\sigma\rangle - j^*(H,\sigma)\Big\} = \inf_{\zeta\in\mathcal{S}^{d\times d}_+} \max_{\sigma\in\mathcal{S}^{d\times d}} \Big\{\langle \xi+\zeta,\sigma\rangle - j^*(H,\sigma)\Big\} \\ &= \inf_{\zeta\in\mathcal{S}^{d\times d}_+} j(H,\xi+\zeta), \end{aligned} \tag{3.37}$$

where, thanks to the coerciveness showed above, in each maximization problem the solution exists while in order to swap the order of inf and max we may use [80, Chapter VI, Proposition 2.3]. Formula (3.36) is proved.

Conditions (A1), (A2) for j_+ follow directly from definitions (3.35) and properties of the convex conjugate. Condition (A3) can be inferred from (3.36), where function $j_+(\,\cdot\,,\xi)$ is a

point-wise infimum of concave upper semi-continuous functions $j(\,\cdot\,, \xi + \zeta)$; one similarly shows (A4).

From the chain of equalities (3.37) we infer that $\hat{\zeta} \in \mathcal{S}^{d\times d}_+$ solves the minimization problem $\inf_{\zeta\in\mathcal{S}^{d\times d}_+} j(H, \xi + \zeta)$ if and only if there exists $\hat{\sigma} \in \mathcal{S}^{d\times d}$ such that $(\hat{\zeta}, \hat{\sigma})$ is a saddle point for the functional $(\zeta, \sigma) \mapsto \langle \xi + \zeta, \sigma\rangle - j^*(H, \sigma)$, i.e. if and only if

$$\max_{\sigma\in\mathcal{S}^{d\times d}} \Big\{ \Big\langle \xi + \hat{\zeta}, \sigma \Big\rangle - j^*(H, \sigma) \Big\} = \langle \xi + \hat{\zeta}, \hat{\sigma}\rangle - j^*(H, \hat{\sigma}) = \min_{\zeta\in\mathcal{S}^{d\times d}_+} \Big\{ \langle \xi + \zeta, \hat{\sigma}\rangle - j^*(H, \hat{\sigma}) \Big\},$$

while the left equation is equivalent to $j(H, \xi+\hat{\zeta}) = \langle \xi + \hat{\zeta}, \hat{\sigma}\rangle - j^*(H, \hat{\sigma})$, i.e. $\hat{\sigma} \in \partial j(H, \xi+\hat{\zeta})$ and the right one to a pair of conditions $\sigma \in \mathcal{S}^{d\times d}_+$ and $\langle \hat{\zeta}, \hat{\sigma}\rangle = 0$. In addition every such $\hat{\sigma}$ solves the first maximization problem in (3.37) hence necessarily $\hat{\sigma} \in j_+(H, \xi)$. □

From the proposition above we can see that if the stress σ is positive semi-definite (resp. negative semi-definite) it is enough to check the constitutive law for j in order to assure that the one for j_+ (resp. j_-) is satisfied:

Corollary 3.21. *For $H \in \mathcal{K}$ and $\xi, \sigma \in \mathcal{S}^{d\times d}$ assume that $\sigma \in \mathcal{S}^{d\times d}_+$ and $\sigma \in \partial j(H, \xi)$. Then $\sigma \in \partial j_+(H, \xi)$ as well. Analogous result holds true for $\sigma \in \mathcal{S}^{d\times d}_-$ and j_-.*

Proof. According to the "moreover part" of Proposition 3.20 the zero tensor $\hat{\zeta} = 0$ solves (3.36). □

In general the minimization problem (3.36) is not guaranteed to admit a solution due to the possible lack of coerciveness of $j(H, \,\cdot\,)$. If the latter is enforced we can characterize the Hooke's law $\sigma \in \partial j_+(H, \sigma)$ in terms of the starting elastic potential $j(H, \,\cdot\,)$:

Corollary 3.22. *For chosen $H \in \mathcal{K}$ let us assume that $j(H, \,\cdot\,) \geq C|\,\cdot\,|^p$ on $\mathcal{S}^{d\times d}$ for some positive constant $C > 0$. Then, for any pair $\xi, \sigma \in \mathcal{S}^{d\times d}$ the following conditions are equivalent:*

(i) the constitutive law that rules out compression holds:

$$\sigma \in \partial j_+(H, \xi);$$

(ii) σ is an element of $\mathcal{S}^{d\times d}_+$ and there exists $\hat{\zeta} \in \mathcal{S}^{d\times d}_+$ such that

$$\sigma \in \partial j(H, \xi + \hat{\zeta}) \qquad \text{and} \qquad \langle \hat{\zeta}, \sigma\rangle = 0.$$

The analogous statement holds for j_- provided that above $\mathcal{S}^{d\times d}_+$ is replaced with $\mathcal{S}^{d\times d}_-$.

Proof. Implication (ii) ⇒ (i) follows directly from Proposition 3.20. Conversely, assuming (i), characterization (3.5) gives

$$\langle \xi, \sigma\rangle = j_+(H, \xi) + j^*_+(H, \sigma) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+} j(H, \xi + \zeta) + j^*(H, \sigma) + \chi_{\mathcal{S}^{d\times d}_+}(\sigma)$$

which immediately implies that $\sigma \in \mathcal{S}^{d\times d}_+$. The assumption yields coerciveness of $j(H, \,\cdot\,)$ therefore solution $\zeta = \hat{\zeta} \in \mathcal{S}^{d\times d}_+$ of minimization problem above exists. As a result

$$\langle \xi + \hat{\zeta}, \sigma\rangle = \langle \xi, \sigma\rangle + \langle \hat{\zeta}, \sigma\rangle \geq \langle \xi, \sigma\rangle = j(H, \xi + \hat{\zeta}) + j^*(H, \sigma)$$

while Fenchel's inequality reads: $\langle \xi + \hat{\zeta}, \sigma\rangle \leq j(H, \xi + \hat{\zeta}) + j^*(H, \sigma)$. Hence, we have equalities above and (ii) holds true by (3.5). The proof for j_- is analogous. □

It is clear that the assumption ($\mathrm{A}_{\mathrm{ellip}}$) is not satisfied for either of functions j_+, j_- even if it is satisfied for the original j. Indeed, from (3.36) we see that for any $\xi \in \mathcal{S}^{d\times d}_-$ we have $j_+(H,\xi) \le j\big(H, \xi + (-\xi)\big) = 0$. On the other hand, if $\xi \notin \mathcal{S}^{d\times d}_-$ clearly there is no $\zeta \in \mathcal{S}^{d\times d}_+$ such that $\xi + \zeta = 0$. We thus obtain:

Proposition 3.23. *Let j satisfy the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$). Then $\bar{j}_+(\xi) = 0$ (resp. $\bar{j}_-(\xi) = 0$) if and only if $\xi \in \mathcal{S}^{d\times d}_-$ (resp. $\xi \in \mathcal{S}^{d\times d}_+$).*

Example 3.24 (**Constitutive law of isotropic masonry material**)**.** We use the newly derived functions to recast the material model for 2D masonry structures proposed in [90]; note that therein authors depart from the material's model similar to (ii) in Corollary 3.22. For $\mathcal{K} = \mathcal{K}_{iso}$ being the cone of isotropic Hooke tensors we take the classical energy function $j(H,\xi) = \frac{1}{2}\langle H\xi, \xi\rangle = \frac{1}{2}\left(\mathscr{K}\,|\mathrm{Tr}\,\xi|^2 + 2\mathscr{G}\,|\mathrm{dev}\,\xi|^2\right)$. We propose the elastic energy function for masonry material:

$$j_-(H,\xi) = \inf_{\zeta\in\mathcal{S}^{2\times 2}_-} j(H,\xi+\zeta). \tag{3.38}$$

For a fixed tensor ξ we shall compute the explicit formula for $j_-(H,\xi)$ by finding the optimal ζ via the optimality criteria from Proposition 3.11. In what follows we will denote $\varepsilon_i := \lambda_i(\xi)$ with $\varepsilon_1 \le \varepsilon_2$ while v_1, v_2 will stand for eigenvectors of ξ, i.e. $\xi = \sum_{i=1}^2 \varepsilon_i\, v_i \otimes v_i$.

First we clear up the easiest case: if $\varepsilon_1, \varepsilon_2 \ge 0$, i.e. $\xi \in \mathcal{S}^{2\times 2}_+$ then $j_-(H,\xi) = 0$, see Proposition 3.23. Further we thus assume that $\varepsilon_1 < 0$. We propose to look at $\zeta = z_2\, v_2 \otimes v_2 \in \mathcal{S}^{2\times 2}_-$, i.e. $z_2 \le 0$. For $H \in \mathcal{K}_{iso}$ straightforward computations show that the constitutive law $\sigma \in j(H, \xi+\zeta)$ furnishes

$$\sigma = H(\xi+\zeta) \quad\Leftrightarrow\quad \sigma = \sum_{i=1}^2 s_i\, v_i\otimes v_i, \quad \begin{cases} s_1 = \mathscr{K}(\varepsilon_1+\varepsilon_2) + \mathscr{G}(\varepsilon_1-\varepsilon_2) + (\mathscr{K}-\mathscr{G})z_2, \\ s_2 = \mathscr{K}(\varepsilon_1+\varepsilon_2) + \mathscr{G}(\varepsilon_2-\varepsilon_1) + (\mathscr{K}+\mathscr{G})z_2. \end{cases}$$

First we look at the singular case when $\mathscr{K} = 0$ for which we have $s_1 = -s_2$. We choose $z_2 = \varepsilon_1 - \varepsilon_2 \le 0$ for which we obtain $\sigma = 0$ and clearly $\langle\zeta,\sigma\rangle = 0$. Therefore by Proposition 3.11 tensor ζ is a solution for (3.38), which establishes $j_+(H,\xi) = j(H,\xi+\zeta) = \frac{1}{2}\langle H(\xi+\zeta),(\xi+\zeta)\rangle = \frac{1}{2}\langle\sigma,(\xi+\zeta)\rangle = 0$.

Further we shall thus assume that $\mathscr{K} > 0$. Since $\varepsilon_1 \le \varepsilon_2$ we see that $s_1 \le s_2$ for $z_2 = 0$, therefore for $z_2 = 0$ there holds $\sigma \in \mathcal{S}^{2\times 2}_-$ if and only if $\mathscr{K}(\varepsilon_1+\varepsilon_2) + \mathscr{G}(\varepsilon_2-\varepsilon_1) \le 0$. For any $\xi \notin \mathcal{S}^{2\times 2}_+$ we thus propose:

$$\zeta = z_2\, v_2 \otimes v_2, \qquad z_2 = -\frac{\big(\mathscr{K}(\varepsilon_1+\varepsilon_2) + \mathscr{G}(\varepsilon_2-\varepsilon_1)\big)_+}{\mathscr{K}+\mathscr{G}},$$

where $(\,\cdot\,)_+$ stands for the positive part of a number.

Consequently, if $z_2 = 0$ (which, in particular, is the case for $\xi \in \mathcal{S}^{2\times 2}_-$) then $\sigma = H(\xi+\zeta) = H\xi \in \mathcal{S}^{2\times 2}_-$ while obviously $\langle\zeta,\sigma\rangle = 0$. By Proposition 3.11 tensor $\zeta = 0$ solves the minimization problem and $j_+(H,\xi) = j(H,\xi)$.

If, on the other hand, $z_2 < 0$ then $\sigma \in \mathcal{S}^{2\times 2}_-$ is a rank-one tensor as $s_2 = 0$ and $s_1 \le 0$; in addition $\langle\zeta,\sigma\rangle = z_2\, s_2 = 0$. Again by virtue of Proposition 3.11 tensor ζ is a solution. Elementary computations show

$$j_+(H,\xi) = j(H,\xi+\zeta) = \frac{1}{2}\langle\sigma,(\xi+\zeta)\rangle = \frac{1}{2}s_1\,\varepsilon_1 = \frac{1}{2}\,\frac{4\,\mathscr{K}\mathscr{G}}{\mathscr{K}+\mathscr{G}}\,(\varepsilon_1)^2.$$

We gather the obtained results and give the full formula for j_- below; we recall that $\xi = \sum_{i=1}^{2} \varepsilon_i\, v_i \otimes v_i$ with $\varepsilon_1 \le \varepsilon_2$:

- for $\mathscr{K} > 0$ we have

$$j_-(H,\xi) = \begin{cases} j(H,\xi) & \text{if} \quad \varepsilon_1 + \varepsilon_2 \le \frac{\mathscr{G}}{\mathscr{K}}\,(\varepsilon_1 - \varepsilon_2), \\ \frac{1}{2}\,\frac{4\mathscr{K}\mathscr{G}}{\mathscr{K}+\mathscr{G}}\,(\varepsilon_1)^2 & \text{if} \quad \frac{\mathscr{G}}{\mathscr{K}}\,(\varepsilon_1 - \varepsilon_2) < \varepsilon_1 + \varepsilon_2 < \varepsilon_2 - \varepsilon_1, \\ 0 & \text{if} \quad \xi \in \mathcal{S}^{2\times 2}_+; \end{cases} \tag{3.39}$$

- if $\mathscr{K} = 0$ we have $j_-(H,\xi) = 0$.

After recalling the one-to-one correspondence between the pair of moduli $\mathscr{K}$ and $\mathscr{G}$ and the pair: Young modulus E, Poisson ratio ν

$$E = 2\left(\frac{1}{2\mathscr{K}} + \frac{1}{2\mathscr{G}}\right)^{-1} = \frac{4\mathscr{K}\mathscr{G}}{\mathscr{K}+\mathscr{G}}, \qquad \nu = \frac{\mathscr{K}-\mathscr{G}}{\mathscr{K}+\mathscr{G}} \tag{3.40}$$

we find that the formula for the energy j_- is identical to the formula [90, Eq. (3.19)] which was derived in a more elementary way. △

Remark 3.25. Up to this point every example considered in this thesis referred to the energy function $j = j(H,\xi)$ that was linear in H. We can see that this is not the case for Example 3.24: the energy function $j_-(H,\xi)$ modelling an isotropic masonry material varies in H (i.e. in $\mathscr{K},\mathscr{G}$) non-linearly provided that $\xi \notin \mathcal{S}^{2\times 2}_+$ and we move in the set $\left\{(\mathscr{K},\mathscr{G}) : \frac{\mathscr{G}}{\mathscr{K}}\,(\varepsilon_1 - \varepsilon_2) < \varepsilon_1 + \varepsilon_2 < \varepsilon_2 - \varepsilon_1\right\}$. This example is an important use of a fairly general assumptions (A3), (A4) that we keep throughout the work.

Now that we know that j_+, j_- are valid energy functions we state several facts about optimal design of the anisotropy:

Proposition 3.26. *If ρ, ρ^0 and $\bar{\mathcal{K}}_1(\xi), \underline{\mathcal{K}}_1(\sigma)$ are the gauges and the optimal sets computed for j, then for j_+ the following statements hold:*

(i) The gauge functions satisfying $\bar{j}_+ = \frac{1}{p}(\rho_+(\,\cdot\,))^p$ and $\bar{j}^*_+ = \frac{1}{p'}(\rho^0_+(\,\cdot\,))^{p'}$ read

$$\rho_+(\xi) = \inf_{\zeta \in \mathcal{S}^{d\times d}_+} \rho(\xi+\zeta), \qquad \rho^0_+(\sigma) = \rho^0(\sigma) + \chi_{\mathcal{S}^{d\times d}_+}(\sigma). \tag{3.41}$$

(ii) Assume that $\rho_+(\xi) > 0$. Then, a tensor $\hat{\zeta} \in \mathcal{S}^{d\times d}_+$ is a solution of the above minimization problem for a given ξ if and only if there exists a non-zero $\hat{\sigma} \in \mathcal{S}^{d\times d}_+$ such that

$$\langle \xi + \hat{\zeta}, \hat{\sigma}\rangle = \rho(\xi+\hat{\zeta})\,\rho^0(\hat{\sigma}) \qquad \text{and} \qquad \langle \hat{\zeta}, \hat{\sigma}\rangle = 0,$$

when in addition $\langle \xi, \hat{\sigma}\rangle = \rho_+(\xi)\,\rho^0_+(\hat{\sigma})$ holds. Moreover, existence of the optimal tensor $\hat{\zeta}$ is assured once the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$) is assumed for j.

(iii) The optimal sets of Hooke tensors satisfy:

$$\bar{\mathcal{K}}_{1,+}(\xi) \subset \bar{\mathcal{K}}_1(\xi+\hat{\zeta}), \qquad \underline{\mathcal{K}}_{1,+}(\sigma) = \begin{cases} \underline{\mathcal{K}}_1(\sigma) & \text{if} \quad \sigma \in \mathcal{S}^{d\times d}_+, \\ \mathcal{K}_1 & \text{if} \quad \sigma \notin \mathcal{S}^{d\times d}_+. \end{cases} \tag{3.42}$$

where $\hat{\zeta}$ is any solution of the minimization problem in (3.41).

Analogous statements hold for ρ_-, ρ^0_- and $\bar{\mathcal{K}}_{1,-}(\xi), \underline{\mathcal{K}}_{1,-}(\sigma)$ computed for j_- provided that above $\mathcal{S}^{d\times d}_+$ is replaced by $\mathcal{S}^{d\times d}_-$.

Proof. We start with a remark concerning statement (ii): since $\rho_+(\xi) > 0$ we must have $\rho(\xi + \hat{\zeta}) > 0$ and, because $\hat{\sigma}$ is determined up to a multiplicative constant, we may assume in the sequel of the proof that: $(\rho^0(\hat{\sigma}))^{p'} = (\rho(\xi + \zeta))^p$.

For each $\sigma \in \mathcal{S}^{d\times d}$ we have

$$\begin{aligned}\frac{1}{p'}(\rho^0_+(\sigma))^{p'} = \bar{j}^*_+(\sigma) &= \inf_{H\in\mathcal{K}_1} j^*_+(H,\sigma) = \inf_{H\in\mathcal{K}_1}\big(j^*(H,\sigma) + \chi_{\mathcal{S}^{d\times d}_+}(\sigma)\big)\\ &= \Big(\inf_{H\in\mathcal{K}_1} j^*(H,\sigma)\Big) + \chi_{\mathcal{S}^{d\times d}_+}(\sigma) = \bar{j}^*(\sigma) + \chi_{\mathcal{S}^{d\times d}_+}(\sigma) = \frac{1}{p'}(\rho^0(\sigma))^{p'} + \chi_{\mathcal{S}^{d\times d}_+}(\sigma)\end{aligned}$$

which gives formula for ρ^0_+ in (i) and for the set $\underline{\mathcal{K}}_{1,+}(\sigma)$ in (iii).

In the proof of Proposition 3.20 (up to verifying coerciveness of $j^*(H,\cdot)$) one may notice that dependence of j_+, j^*_+ on H is immaterial. One could therefore repeat the proof with $\bar{j}$, $\bar{j}^*$ in place of $j(H,\cdot)$, $j^*(H,\cdot)$ (coerciveness of $\bar{j}^* = \frac{1}{p}(\rho_+(\cdot))^p$ is even more direct). As a result we obtain $\bar{j}_+(\xi) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+}\bar{j}(\xi+\zeta)$ from which we infer the formula for ρ_+ in (i). Moreover for a fixed ξ we find that $\hat{\zeta}$ solves this minimization problem if and only if there exists $\hat{\sigma} \in \mathcal{S}^{d\times d}_+$ such that $\hat{\sigma} \in \partial\bar{j}(H, \xi+\hat{\zeta})$ and $\langle\hat{\zeta},\hat{\sigma}\rangle = 0$. Since $(\rho^0(\hat{\sigma}))^{p'} = (\rho(\xi+\zeta))^p$ this statement is equivalent to the condition given in (ii) by virtue of Proposition 3.12. Once ($\mathrm{A}_{\mathrm{ellip}}$) is valid the function ρ is coercive an existence of optimal $\hat{\zeta}$ follows.

It is left to show the inclusion $\bar{\mathcal{K}}_{1,+}(\xi) \subset \underline{\mathcal{K}}_1(\xi+\hat{\zeta})$. By Ky Fan's theorem we have:

$$\bar{j}_+(\xi) = \max_{H\in\mathcal{K}_1} j_+(H,\xi) = \max_{H\in\mathcal{K}_1}\inf_{\zeta\in\mathcal{S}^{d\times d}_+} j(H,\xi+\zeta) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+}\max_{H\in\mathcal{K}_1} j(H,\xi+\zeta) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+}\bar{j}(\xi+\zeta).$$

If we take a solution $\hat{\zeta}$ of the last problem in the chain and a solution $\hat{H}$ of the first problem, i.e. $\hat{H} \in \bar{\mathcal{K}}_{1,+}(\xi)$, then $(\hat{\zeta},\hat{H})$ is a saddle point of the functional $(\zeta,H)\mapsto j(H,\xi+\zeta)$ and in particular $\hat{H}$ solves $\max_{H\in\mathcal{K}_1} j(H,\xi+\hat{\zeta})$, i.e. $\hat{H} \in \bar{\mathcal{K}}_1(\xi+\hat{\zeta})$. □

Example 3.27 (**Isotropic Masonry Material Design problem**)**.** We put forward the two dimensional ($d=2$) isotropic *Masonry Material Design* problem (MMD) where:

$$\mathcal{K} = \mathcal{K}_{iso}, \qquad j_-(H,\xi) \ \text{ for } \ j(H,\xi) = \frac{1}{2}\langle H\xi,\xi\rangle, \qquad c(H) = \operatorname{Tr} H.$$

We shall compute ρ_-, ρ^0_- and $\underline{\mathcal{K}}_{1,-}(\sigma)$ by employing Proposition 3.26. Immediately we obtain that for any $\sigma \in \mathcal{S}^{2\times 2}_-$ we have

$$\begin{aligned}\rho^0_-(\sigma) = \rho^0(\sigma) &= \frac{1}{\sqrt{2}}\,|\operatorname{Tr}\sigma| + \sqrt{2}\,|\operatorname{dev}\sigma|,\\ \underline{\mathcal{K}}_{1,-}(\sigma) = \underline{\mathcal{K}}_1(\sigma) &= \left\{H\in\mathcal{K}_{iso} \ :\ \mathscr{K} = \frac{1}{2\sqrt{2}}\,\frac{|\operatorname{Tr}\sigma|}{\rho^0(\sigma)},\ \mathscr{G} = \frac{1}{2\sqrt{2}}\,\frac{|\operatorname{dev}\sigma|}{\rho^0(\sigma)}\right\},\end{aligned}$$

whilst for $\sigma \notin \mathcal{S}^{2\times 2}_-$ we have $\rho^0_-(\sigma) = \infty$ and $\underline{\mathcal{K}}_{1,-}(\sigma) = \mathcal{K}$.

To compute ρ_- we will proceed very similarly to Example 3.24: for a fixed $\xi \in \mathcal{S}^{d\times d}$ we shall find ζ that is optimal for the problem $\rho_-(\xi) = \min_{\zeta\in\mathcal{S}^{d\times d}_-}\rho(\xi+\zeta)$ (that is guaranteed to exist) and to that aim we shall utilize the optimality condition (ii) in Proposition 3.26.

We start with the case when $\xi \in \mathcal{S}^{2\times 2}_+$ for which automatically $\rho_-(\xi) = 0$. Further $\xi \notin \mathcal{S}^{2\times 2}_+$ will be assumed, for which $\xi = \sum_{i=1}^2 \varepsilon_i\, v_i \otimes v_i$ with $\varepsilon_1 \le \varepsilon_2$, $\varepsilon_1 < 0$. Once again it will suffice to limit the search to $\zeta \in \mathcal{S}^{2\times 2}_-$ of the form $\zeta = z_2\, v_2 \otimes v_2 \in \mathcal{S}^{2\times 2}_-$, where $z_2 \le 0$. We compute: $\mathrm{Tr}\,(\xi+\zeta) = \varepsilon_1 + \varepsilon_2 + z_2$ and $|\mathrm{dev}\,(\xi+\zeta)| = \frac{1}{\sqrt{2}}\,|\varepsilon_1 - (\varepsilon_2 + z_2)|$. According to Proposition 3.26 tensor ζ will be a solution provided we find non-zero $\sigma \in \mathcal{S}^{2\times 2}_-$ such that $\langle \zeta, \sigma\rangle = 0$ and the extremality equality $\langle \xi+\zeta, \sigma\rangle = \rho(\xi+\zeta)\,\rho^0(\sigma)$ is met. From the characterization (3.22) we may deduce that eigenvectors of σ and $\xi+\zeta$ must coincide so necessarily $\sigma = \sum_{i=1}^2 s_i\, v_i \otimes v_i$. Upon denoting $\tilde{\varepsilon}_2 := \varepsilon_2 + z_2$ it can be showed that extremality equality holds if and only if

$$\begin{cases} s_1 + s_2 \le 0, \quad s_1 - s_2 = 0 & \text{if} \quad \varepsilon_1 + \tilde{\varepsilon}_2 < -\frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2|, \\ s_1 + s_2 \le 0, \quad (s_1 - s_2)(\varepsilon_1 - \tilde{\varepsilon}_2) \ge 0 & \text{if} \quad \varepsilon_1 + \tilde{\varepsilon}_2 = -\frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2|, \\ s_1 + s_2 = 0, \quad (s_1 - s_2)(\varepsilon_1 - \tilde{\varepsilon}_2) \ge 0 & \text{if} \quad -\frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2| < \varepsilon_1 + \tilde{\varepsilon}_2 \le \frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2|, \\ s_1 + s_2 \ge 0, \quad (s_1 - s_2)(\varepsilon_1 - \tilde{\varepsilon}_2) \ge 0 & \text{if} \quad \varepsilon_1 + \tilde{\varepsilon}_2 = \frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2|, \\ s_1 + s_2 \ge 0, \quad s_1 - s_2 = 0 & \text{if} \quad \frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2| < \varepsilon_1 + \tilde{\varepsilon}_2. \end{cases}$$

In order to ensure that $\sigma \in \mathcal{S}^{2\times 2}_-$, $\sigma \neq 0$ tensor ζ must be chosen so that one of the two top cases occur. First we consider the case when ξ satisfies $\varepsilon_1 + \varepsilon_2 \le -\frac{1}{\sqrt{2}}|\varepsilon_1 - \varepsilon_2|$ (we recall that $\varepsilon_1 - \varepsilon_2 \le 0$). We will show that necessarily $z_2 = 0$. Assume the contrary, i.e. that $z_2 < 0$, then it is easy to check that $\varepsilon_1 + \tilde{\varepsilon}_2 < -\frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2|$ hence we are obliged to choose $s_1 = s_2 < 0$. But then $\langle \sigma, \zeta\rangle = z_2\, s_2 > 0$ so ζ cannot be a solution. For such ξ tensor $\zeta = 0$ is a solution and $\rho_+(\xi) = \rho(\xi)$.

Ultimately, for $\xi \notin \mathcal{S}^{2\times 2}_+$ we choose

$$\zeta = z_2\, v_2 \otimes v_2, \qquad z_2 = -\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\,\varepsilon_1 + \varepsilon_2\right)_+$$

If $z_2 < 0$ we readily check that $\varepsilon_1 + \tilde{\varepsilon}_2 = -\frac{1}{\sqrt{2}}|\varepsilon_1 - \tilde{\varepsilon}_2|$ and then we may choose $s_1 < 0$ and $s_2 = 0$, which gives $\langle \zeta, \sigma\rangle = z_2\, s_2 = 0$ and validates ζ as a solution. If $z_2 < 0$ we compute $\mathrm{Tr}(\xi+\zeta) = -|\mathrm{dev}(\xi+\zeta)| = \frac{2}{\sqrt{2}+1}\,\varepsilon_1$. As a result $\rho_-(\xi) = \rho(\xi+\zeta) = \max\big\{|\mathrm{Tr}(\xi+\zeta)|/\sqrt{2},\, |\mathrm{dev}(\xi+\zeta)|/\sqrt{2}\big\} = \frac{2}{2+\sqrt{2}}\,|\varepsilon_1|$.

We summarize the results and give the full formula for ρ_- below; we recall that $\xi = \sum_{i=1}^2 \varepsilon_i\, v_i \otimes v_i$ with $\varepsilon_1 \le \varepsilon_2$:

$$\rho_-(\xi) = \begin{cases} \rho(\xi) & \text{if} \quad \varepsilon_1 + \varepsilon_2 \le \frac{1}{\sqrt{2}}\,(\varepsilon_1 - \varepsilon_2), \\ \frac{2}{2+\sqrt{2}}\,|\varepsilon_1| & \text{if} \quad \frac{1}{\sqrt{2}}\,(\varepsilon_1 - \varepsilon_2) < \varepsilon_1 + \varepsilon_2 < \varepsilon_2 - \varepsilon_1, \\ 0 & \text{if} \quad \xi \in \mathcal{S}^{2\times 2}_+. \end{cases} \tag{3.43}$$

We conclude the example with a remark: for the Hooke tensor $\bar{H} \in \mathcal{K}_1$ defined by moduli $\bar{\mathscr{K}} = \frac{1}{2+2\sqrt{2}}$, $\bar{\mathscr{G}} = \frac{1}{4+2\sqrt{2}}$ we have:

$$E = \frac{4\bar{\mathscr{K}}\bar{\mathscr{G}}}{\bar{\mathscr{K}} + \bar{\mathscr{G}}} = \left(\frac{2}{2+\sqrt{2}}\right)^2, \qquad \frac{\bar{\mathscr{G}}}{\bar{\mathscr{K}}} = \frac{1}{\sqrt{2}}.$$

By comparing formulas (3.39) and (3.43) for j_- and, respectively, ρ_- we find that

$$\frac{1}{2}\big(\rho_-(\xi)\big)^2 = j_-(\bar{H}, \xi) \qquad \text{whenever} \quad \frac{1}{\sqrt{2}}\,(\varepsilon_1 - \varepsilon_2) < \varepsilon_1 + \varepsilon_2,$$

i.e. $\bar{H} \in \bar{\mathcal{K}}_{1,-}(\xi)$. In Example 4.34, Case c) we shall learn that $\bar{H} \in \underline{\mathcal{K}}_1(\sigma)$ (for the IMD setting) whenever σ is rank-one and the regime $\frac{1}{\sqrt{2}}(\varepsilon_1 - \varepsilon_2) < \varepsilon_1 + \varepsilon_2$ is precisely the one where the associated σ is rank-one which gives the intuition behind the result. $\triangle$

Although mathematically fully analogous to j_-, the energy function of type j_+ receives a quite different interpretation in terms of mechanics. While j_- modelled a material that is perfectly brittle, that cracks under the smallest tensile stress, the present author's proposal is to employ function j_+ to model a material that is *perfectly immune to buckling*, see Section 5.2 for more insightful explanation. Such a model may be deemed reasonable when considering a 2D sheet of a very small thickness or a *canvas-like* material which has a capacity to wrinkle or fold (the opposite of cracking). Having said this, we shall limit ourselves to energy functions of class j_+ in Chapter 5 which deals with optimal design of pre-stressed plane membranes. Amongst functions of type j_+ the example given below will be central for this chapter and in fact it will be a foundation for Chapter 6 on optimal design of form-structures:

Example 3.28 (**Canvas-like Material Design problem**)**.** We propose the optimal anisotropy design problem where we take:

$$\mathcal{K} = \mathcal{K}_{fib}, \qquad j_+(H,\xi) \ \text{ for } \ j(H,\xi) = \frac{1}{2}\langle H\xi, \xi\rangle, \qquad c(H) = \operatorname{Tr} H.$$

The setting proposed will be referred to as *Canvas-like Material Design* problem (CanMD). The choice of the set of admissible Hooke tensors $\mathcal{K}_{fib}$ ought to reflect the fibrous character of a canvas.

We shall compute ρ_+, ρ_+^0 and the set $\underline{\mathcal{K}}_{1,+}(\sigma)$ again by employing Proposition 3.26 and by utilizing Example 3.18 on the FibMD where also $\mathcal{K} = \mathcal{K}_{fib}$ was chosen. Immediately we obtain for any $\sigma \in \mathcal{S}^{d\times d}_+$

$$\rho_+^0(\sigma) = \rho^0(\sigma) = \sum_{i=1}^d |\lambda_i(\sigma)| = \sum_{i=1}^d \lambda_i(\sigma) = \operatorname{Tr}\sigma.$$

For any $\xi \in \mathcal{S}^{d\times d}$ by $\xi_+, \xi_- \in \mathcal{S}^{d\times d}_+$ we will further denote its positive and negative part: for the spectral decomposition $\xi = \sum_{i=1}^d \lambda_i(\xi)\, v_i \otimes v_i$ we set

$$\xi_+ := \sum_{i=1}^d \big(\lambda_i(\xi)\big)_+ v_i \otimes v_i, \qquad \xi_- := \sum_{i=1}^d \big(\lambda_i(\xi)\big)_- v_i \otimes v_i, \tag{3.44}$$

where for $a \in \mathbb{R}$ symbols $(a)_+, (a)_- \in \mathbb{R}_+$ stand for the positive and, respectively, negative part of a, i.e. $(a)_+ = \max\{a, 0\}$, $(a)_- = \max\{-a, 0\}$. Clearly $\xi = \xi_+ - \xi_-$ and $\langle \xi_+, \xi_-\rangle = 0$. It is straightforward to check that

$$\rho_+(\xi) = \min_{\zeta \in \mathcal{S}^{d\times d}_+} \rho(\xi + \zeta) = \rho\big(\xi + \xi_-\big) = \rho\big(\xi_+\big) = \max_{i\in\{1,\dots,d\}} \big(\lambda_i(\xi)\big)_+,$$

i.e. that $\zeta = \xi_-$ solves the minimization problem above. The extremality conditions are an easy modification of conditions (3.29) for the FibMD setting: for any non-zero $\sigma \in \mathcal{S}^{d\times d}$ and $\xi \in \mathcal{S}^{d\times d}$ such that $\rho_+(\xi) > 0$ we have the equivalence

$$\langle \xi, \sigma\rangle = \rho_+(\xi)\, \rho_+^0(\sigma) \quad \Leftrightarrow \quad \begin{cases} \text{every eigenvector of } \sigma \text{ is an eigenvector of } \xi \text{ and} \\ \lambda_i(\sigma) = \frac{(\lambda_i(\xi))_+}{\rho_+(\xi)}\, |\lambda_i(\sigma)|. \end{cases}$$

We note that conditions on the RHS imply that $\sigma \in \mathcal{S}^{d\times d}_+$.

Based on point (iii) in Proposition 3.26 we have $\underline{\mathcal{K}}_{1,+}(\sigma) = \underline{\mathcal{K}}_1(\sigma)$ for any $\sigma \in \mathcal{S}^{d\times d}_+$. The set $\underline{\mathcal{K}}_1(\sigma)$ has been already characterized in Example 3.18, yet only for $d = 2$ and here we are aiming for full generality as we shall use the CanMD setting both for $d = 2$ and $d = 3$. In a way, the steps taken below will be a combination of arguments used in Case b) and Case c) in Example 3.18.

For any non-zero $\sigma \in \mathcal{S}^{d\times d}_+$ we will compute $\underline{\mathcal{K}}_1(\sigma)$ (for problem FibMD instead of CanMD directly) via the optimality criteria in point (ii) of Theorem 3.14. We choose $\xi_\sigma = P_{\operatorname{Im}\sigma}$, i.e. the orthogonal projection onto image of σ. Then, every $\tau \in \operatorname{Im}\sigma \cap S^{d-1}$ is an eigenvector of ξ_σ with the corresponding eigenvalue equal to 1 and we see that the extremality equality $\langle \xi, \sigma\rangle = \rho(\xi)\,\rho^0(\sigma)$ holds, cf. (3.29). By (3.27) we have $\bar{\mathcal{K}}_1(\xi_\sigma) = \operatorname{co}\big\{\tau\otimes\tau\otimes\tau\otimes\tau \,:\, \tau \in \operatorname{Im}\sigma\cap S^{d-1}\big\}$ and owing to point (iii) of Theorem 3.14 we have $\underline{\mathcal{K}}_1(\sigma) \subset \bar{\mathcal{K}}_1(\xi_\sigma)$, namely we limit the search to tensors of the form $H = \sum_{i=1}^m \alpha_i\,\tau_i\otimes\tau_i\otimes\tau_i\otimes\tau_i$ for some $m \in \mathbb{N}_+$, $\tau_i \in \operatorname{Im}\sigma\cap S^{d-1}$, $\alpha_i \geq 0$ and $\sum_{i=1}^m \alpha_i = 1$. Eventually, by virtue of point (ii) of Theorem 3.14, H of this form is an element of $\underline{\mathcal{K}}_1(\sigma)$ if and only if the constitutive law $\sigma/\rho^0(\sigma) \in \partial j(H, \xi_\sigma)$ holds, i.e. $\sigma/(\operatorname{Tr}\sigma) = H\xi_\sigma$. Upon recalling that $\underline{\mathcal{K}}_{1,+}(\sigma) = \underline{\mathcal{K}}_1(\sigma)$ whenever $\sigma \in \mathcal{S}^{d\times d}_+$ we arrive at the characterization (recall that $\langle \tau_i\otimes\tau_i, \xi_\sigma\rangle = 1 \;\; \forall\, i$):

$$\underline{\mathcal{K}}_{1,+}(\sigma) = \left\{ \sum_{i=1}^m \alpha_i\,\tau_i\otimes\tau_i\otimes\tau_i\otimes\tau_i \,:\, \tau_i \in \operatorname{Im}\sigma\cap S^{d-1},\; \sum_{i=1}^m \alpha_i = 1,\; \frac{\sigma}{\operatorname{Tr}\sigma} = \sum_{i=1}^m \alpha_i\,\tau_i\otimes\tau_i \right\}.$$

A particular optimal Hooke tensor may be written down as

$$\underline{H}_\sigma = \sum_{i=1}^d \frac{\lambda_i(\sigma)}{\operatorname{Tr}\sigma}\; v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma) \quad \in \quad \underline{\mathcal{K}}_{1,+}(\sigma), \tag{3.45}$$

where $v_i(\sigma)$ are the eigenvectors of σ. The Hooke tensor $\underline{H}_\sigma$ may be viewed as a composition of a family of orthogonal material fibres. △

Remark 3.29. In the CanMD setting, when σ is of rank higher than one, the set $\underline{\mathcal{K}}_{1,+}(\sigma)$ turns out to contain "many" Hooke tensors that are non-obvious. For $d = 2$ let $\mathrm{e}_1, \mathrm{e}_2$ be a Cartesian base of $\mathbb{R}^2$, we consider

$$\sigma = \frac{4}{5}\,\mathrm{e}_1\otimes\mathrm{e}_1 + \frac{1}{5}\,\mathrm{e}_2\otimes\mathrm{e}_2 \in \mathcal{S}^{2\times 2}_+;$$

we see that $\operatorname{Tr}\sigma = 1$. According to (3.45) one of the optimal Hooke tensors reads

$$\underline{H}_\sigma = \frac{4}{5}\,\mathrm{e}_1\otimes\mathrm{e}_1\otimes\mathrm{e}_1\otimes\mathrm{e}_1 + \frac{1}{5}\,\mathrm{e}_2\otimes\mathrm{e}_2\otimes\mathrm{e}_2\otimes\mathrm{e}_2 \quad \in \quad \underline{\mathcal{K}}_{1,+}(\sigma).$$

At the same time, by exploiting the characterization of $\underline{\mathcal{K}}_{1,+}(\sigma)$ it is easy to check that the Hooke tensor

$$\tilde{\underline{H}}_\sigma = \sum_{i=1}^2 \frac{1}{2}\,\tau_i\otimes\tau_i\otimes\tau_i\otimes\tau_i, \qquad \tau_1 = \frac{2}{\sqrt{5}}\,\mathrm{e}_1 + \frac{1}{\sqrt{5}}\,\mathrm{e}_2, \quad \tau_2 = \frac{2}{\sqrt{5}}\,\mathrm{e}_1 - \frac{1}{\sqrt{5}}\,\mathrm{e}_2$$

is an element of $\underline{\mathcal{K}}_{1,+}(\sigma)$ as well. Of course $\underline{H}_\sigma \neq \tilde{\underline{H}}_\sigma$. We can readily see that there are infinitely many possibilities of constructing optimal Hooke tensors by superposing families of uni-axial Hooke tensors running at arbitrary angles with respect to $\mathrm{e}_1, \mathrm{e}_2$; note that

any convex combination of $\underline{H}_\sigma$ and $\tilde{H}_\sigma$ is an element of $\mathcal{K}_{1,+}(\sigma)$ as well. This would be intuitive for a hydrostatic stress $\sigma = s\,\mathrm{I}$ whilst, in author's opinion, it is unexpected for σ with $\lambda_1(\sigma) \neq \lambda_2(\sigma)$. This phenomenon will be reflected in the numerical simulations performed in Chapters 5 and 6, see Example 5.73.

The classes of functions j_+ and j_- describe extreme cases of materials that are not able to withstand compression and tension, respectively. They are singular in the sense that the assumption ($\mathrm{A}_{\mathrm{ellip}}$) does not hold. Another possibility is to model a material that behaves dissymmetrically in tension and compression, e.g. concrete (in the phase prior to cracking). For j_+ and j_- defined as in (3.35) and positive reals $\kappa_+, \kappa_- > 0$ we propose an energy function:

$$j_\pm(H,\xi) := (\kappa_+)^p\, j_+(H,\xi) + (\kappa_-)^p\, j_-(H,\xi);$$

raising κ_+, κ_- to the power p will prove convenient in due course. Let us stress that $\pm$ is an independent symbol that emphasizes dissymmetry in tension and compression. By employing the inf-convolution formula for convex conjugate of sum of functions we find that

$$j^*_\pm(H,\sigma) = \inf_{\substack{\sigma_+\in\mathcal{S}^{d\times d}_+\\ \sigma_-\in\mathcal{S}^{d\times d}_-}} \left\{ \frac{1}{(\kappa_+)^{p'}}\, j^*(H,\sigma_+) + \frac{1}{(\kappa_-)^{p'}}\, j^*(H,\sigma_-) \,:\, \sigma_+ + \sigma_- = \sigma \right\}. \tag{3.46}$$

By combining the results in Propositions 3.20 and 3.23 we immediately obtain:

Proposition 3.30. *Let j be an energy function satisfying the assumptions* (A1)-(A4). *Then $j_\pm$ defined above also satisfies* (A1)-(A4). *Moreover $j_\pm$ satisfies* ($\mathrm{A}_{\mathrm{ellip}}$) *if and only if j satisfies* ($\mathrm{A}_{\mathrm{ellip}}$).

Remark 3.31. The function $j_\pm$ could be constructed by using $j_{1,+}$ and $j_{2,-}$ where j_1, and j_2 are any two energy functions satisfying (A1)-(A4). This makes it possible to introduce even more contrast in material's behaviour under tension and compression.

Example 3.32 (**Fibrous Material Design problem with dissymmetry in tension and compression**)**.** We revisit the problem of Fibrous Material Design with the linear constitutive law replaced by the constitutive law for material that responds differently in tension and compression (the design problem will be further abbreviated by FibMD±), i.e. for $\kappa_+, \kappa_- > 0$ we take

$$\mathcal{K} = \mathcal{K}_{fib}, \quad j_\pm(H,\xi) = (\kappa_+)^2\, j_+(H,\xi) + (\kappa_-)^2\, j_-(H,\xi), \quad c(H) = \operatorname{Tr} H \tag{3.47}$$

where j_+, j_- are computed for $j(H,\xi) = \frac{1}{2}\langle H\,\xi,\xi\rangle$. In contrast to Example 3.18 we have no linearity of $j_\pm$ with respect to H and therefore for given $\xi \in \mathcal{S}^{d\times d}$ we must test $j_\pm(H,\xi)$ with tensors H in the whole $\mathcal{K}_{fib}$ instead of just $\mathcal{K}_{axial}$. Below by $\xi_+, \xi_- \in \mathcal{S}^{d\times d}_+$ we will denote the positive and negative part of $\xi = \xi_+ - \xi_-$, see (3.44). Then, for any $H \in \mathcal{K}_1$, i.e. for $H = \sum_{i=1}^m \alpha_i\, \tau_i\otimes\tau_i\otimes\tau_i\otimes\tau_i$ with $\sum_{i=1}^m \alpha_i = 1$, we estimate

$$j_+(H,\xi) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+} \left\{ \sum_{i=1}^m \frac{\alpha_i}{2} \big(\langle \xi+\zeta, \tau_i\otimes\tau_i\rangle\big)^2 \right\} \le \sum_{i=1}^m \frac{\alpha_i}{2} \big(\langle \xi_+, \tau_i\otimes\tau_i\rangle\big)^2$$

and by repeating an analogous estimate for j_- we obtain

$$j_\pm(H,\xi) \le \sum_{i=1}^m \frac{\alpha_i}{2}\left(\big(\kappa_+\langle\xi_+,\tau_i\otimes\tau_i\rangle\big)^2 + \big(\kappa_-\langle-\xi_-,\tau_i\otimes\tau_i\rangle\big)^2\right)$$
$$\le \sum_{i=1}^m \frac{\alpha_i}{2}\left(\big\langle\big(\kappa_+\xi_+ + \kappa_-\xi_-\big),\tau_i\otimes\tau_i\big\rangle\right)^2$$

where we used the fact that $\big(\kappa_+\langle\xi_+,\tau_i\otimes\tau_i\rangle\big)\ge 0$, $\big(\kappa_-\langle\xi_-,\tau_i\otimes\tau_i\rangle\big)\ge 0$. We see that for any $H\in\mathcal{K}_1$ we have

$$j_\pm(H,\xi) \le \sum_{i=1}^m \frac{\alpha_i}{2}\big(\rho_\pm(\xi)\big)^2 \le \frac{1}{2}\big(\rho_\pm(\xi)\big)^2$$

where, upon denoting by ρ the spectral norm from (3.28), we introduce

$$\rho_\pm(\xi) := \rho\big(\kappa_+\,\xi_+ + \kappa_-\,\xi_-\big) = \max_{i\in\{1,\dots,d\}}\left\{\max\big\{\kappa_+\lambda_i(\xi),-\kappa_-\lambda_i(\xi)\big\}\right\}.$$

By choosing τ parallel to a suitable eigenvector of ξ we easily obtain $\bar{j}_\pm(\xi)\ge j_\pm(\tau\otimes\tau\otimes\tau\otimes\tau,\xi) = \frac{1}{2}\big(\rho_\pm(\xi)\big)^2$. These two estimates furnish $\bar{j}_\pm(\xi) = \frac{1}{2}\big(\rho_\pm(\xi)\big)^2$ hence $\rho_\pm$ is the gauge function for the FibMD$\pm$ problem. We observe that

$$\rho_\pm(\xi)\le 1 \qquad\Leftrightarrow\qquad -\frac{1}{\kappa_-}\le\lambda_i(\xi)\le\frac{1}{\kappa_+}\quad\forall\, i\in\{1,\dots,d\}.$$

For $\sigma\in\mathcal{S}^{d\times d}$ the polar $\rho^0_\pm$ reads

$$\rho^0_\pm(\sigma) = \sum_{i=1}^d \max\left\{\frac{\lambda_i(\sigma)}{\kappa_+},-\frac{\lambda_i(\sigma)}{\kappa_-}\right\} = \frac{1}{2}\left(\frac{1}{\kappa_+}-\frac{1}{\kappa_-}\right)\operatorname{Tr}\sigma + \frac{1}{2}\left(\frac{1}{\kappa_+}+\frac{1}{\kappa_-}\right)\rho^0(\sigma)$$

where ρ^0 is the polar to the spectral norm, see (3.28); it is worth to note that $\operatorname{Tr}\sigma$ enters the formula with a sign. The extremality conditions between ξ and σ for $\rho_\pm$ and $\rho^0_\pm$ are very similar to those displayed for FibMD problem (see (3.29)) thus we shall not write them down. The same goes for characterizations of the sets $\bar{\mathcal{K}}_{1,\pm}(\xi)$ and $\underline{\mathcal{K}}_{1,\pm}(\sigma)$; we merely show a formula for

$$\underline{H}_\sigma = \sum_{i=1}^d \frac{\max\big\{\frac{1}{\kappa_+}\lambda_i(\sigma),-\frac{1}{\kappa_-}\lambda_i(\sigma)\big\}}{\rho^0_\pm(\sigma)}\; v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma) \tag{3.48}$$

being a universal (but in general non-unique) element of the set $\underline{\mathcal{K}}_{1,\pm}(\sigma)$. △

3.4. Constitutive fields as tensor valued measures and the energy integral functional

For a bounded domain Ω the *constitutive fields* will be modelled by a vector valued (or, in the context of elasticity or conductivity, tensor valued) measures $\lambda\in\mathcal{M}(\overline{\Omega};W)$. With c being the cost norm on W introduced in Section 3.2, according to Radon-Nikodym theorem any measure $\lambda\in\mathcal{M}(\overline{\Omega};W)$ can be represented as follows:

$$\lambda = \mathscr{H}\mu,\qquad \mu\in\mathcal{M}_+(\overline{\Omega}),\quad \mathscr{H}\in L^\infty_\mu\big(\overline{\Omega};W\big),\quad c(\mathscr{H})=1\ \ \mu\text{-a.e.}, \tag{3.49}$$

that is μ can be computed as $c(\lambda)$ (see the Goffman-Serring convention in (2.9)) while $\mathscr{H}$ is the Radon-Nikodym derivative $d\lambda/\big(d\,c(\lambda)\big)$. Unless noted differently, in the sequel of the work by $\lambda = \mathscr{H}\mu$ we shall understand precisely the representation above.

We choose a closed convex cone $\mathcal{K} \subset W$ in accordance with Section 3.1 that restricts the material's anisotropy. By the set of admissible constitutive fields we shall understand

$$\mathcal{M}(\overline{\Omega};\mathcal{K}) := \Big\{\lambda = \mathscr{H}\mu \in \mathcal{M}(\overline{\Omega};W) \,:\, \mathscr{H}(x) \in \mathcal{K} \;\text{ for } \mu\text{-a.e. } x\Big\}.$$

We note that since $\mathcal{K}$ is a cone the definition of $\mathcal{M}(\overline{\Omega};\mathcal{K})$ does not depend on the particular representation $\lambda = \mathscr{H}\mu$, i.e. on the choice of the norm c.

Proposition 3.33. *The set $\mathcal{M}(\overline{\Omega};\mathcal{K})$ is a weakly-* closed convex cone in $\mathcal{M}(\overline{\Omega};W)$.*

Proof. It is enough to notice that $\mathcal{M}(\overline{\Omega};\mathcal{K}) = \big\{\lambda \in \mathcal{M}(\mathbb{R}^d;W) \,:\, F(\lambda) \leq 0\big\}$ where $F(\lambda) = \int_{\mathbb{R}^d} \chi_{\overline{\Omega}}(x)\,\chi_{\mathcal{K}}\big(\mathscr{H}(x)\big)\,c(\lambda)(dx)$. The function $\chi_{\overline{\Omega}}(\,\cdot\,)\,\chi_{\mathcal{K}}(\,\cdot\,) : \mathbb{R}^d \times W \to \overline{\mathbb{R}}$ satisfies the prerequisite of Reshetnyak's theorem and therefore $\mathcal{M}(\overline{\Omega};\mathcal{K})$ is weakly-* closed, cf. Corollary 2.25. □

In (3.49) the information on the constitutive field λ has been split into two: information on the material distribution μ and information on the anisotropy $\mathscr{H}$. The measure μ, whose support may be a proper subset of $\overline{\Omega}$, determines the topology and shape of the body. Additional geometric features of the body may be point-wise identified via the space tangent to measure T_μ (see Section 2.5) – measure λ may account for lower dimensional structural elements such as membrane shells, ribs or bars (in the context of elasticity).

Next, for an energy integral $j : \mathcal{K} \times V \to \mathbb{R}$ that meets the assumptions (A1)-(A4) and for a fixed continuous function $\epsilon \in C(\overline{\Omega};V)$ (strain field in elasticity) we look at the integral functional that for a constitutive field λ furnishes the total energy in the body:

$$\mathcal{M}(\overline{\Omega};\mathcal{K}) \ni \lambda \mapsto \int j\big(\lambda,\epsilon\big) := \int j\big(\mathscr{H}(x),\epsilon(x)\big)\mu(dx) \tag{3.50}$$

where the latter equality ought to be treated as convention similar to the one of Goffman and Serrin, see (2.10). As will unfold in due course the optimal design problems that will be put forward in this thesis will require that the above functional be concave and weakly-* upper semi-continuous. Concavity follows directly from assumption (A3) and below we shall prove the upper semi-continuity result. Beforehand we make a comment:

Remark 3.34. In Section 3.1 we proposed to extend the function $j : \mathcal{K} \times V \to \overline{\mathbb{R}}$ to an upper semi-continuous function on the whole space $W \times V$ by putting $j(H,\xi) = -\infty$ for any $H \in W \setminus \mathcal{K}$. By using the extended function j we automatically extend the functional $\lambda \mapsto \int j\big(\lambda,\epsilon\big)$ to the space $\mathcal{M}(\overline{\Omega};W)$ and this way

$$\int j\big(\lambda,\epsilon\big) = -\infty \qquad \text{for any } \lambda \in \mathcal{M}(\overline{\Omega};W) \setminus \mathcal{M}(\overline{\Omega};\mathcal{K})$$

and ultimately $\mathcal{M}(\overline{\Omega};\mathcal{K}) = \big\{\lambda \in \mathcal{M}(\overline{\Omega};W) : j\big(\lambda,\epsilon\big) > -\infty\big\}$ for arbitrary $\epsilon \in C(\overline{\Omega};V)$.

Instead of proving upper semi-continuity of the concave functional $\lambda \mapsto \int j(\lambda,\epsilon)$ we may focus on the lower semi-continuity of the convex functional $\lambda \mapsto \int j^-(\lambda,\epsilon)$ where simply $j^- := -j$. Then the idea is to use Reshetnyak's theorem (cf. Corollary 2.25),

however, non-positivity of the integrand j^- requires additional trick involving minorization of j^-:

Theorem 3.35. *Let us fix $\epsilon \in C\big(\overline{\Omega}; V\big)$. Then:*

(i) The functional $\lambda \mapsto \int j\big(\lambda, \epsilon\big)$ is concave and weakly- upper semi-continuous on the space $\mathcal{M}\big(\overline{\Omega}; W\big)$.*

(ii) Let $\nu \in \mathcal{M}^b_+(T)$ where T is a locally compact separable metric space while to each t we assign a measure $\lambda^t \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$. We assume that the mapping $t \mapsto \lambda^t$ is ν-measurable and ν-essentially bounded in the sense of Definition 2.26. We define a Radon measure $\lambda_\nu \in \mathcal{M}(\overline{\Omega}; W)$ such that

$$\langle \Psi, \lambda_\nu \rangle_{(C,\mathcal{M})} = \int_T \left(\int_{\overline{\Omega}} \big\langle \Psi, \lambda^t \big\rangle \right) \nu(dt) \qquad \forall\, \Psi \in C(\overline{\Omega}; W).$$

Then λ_ν is an element of $\mathcal{M}(\overline{\Omega}; \mathcal{K})$, i.e. $\frac{d\lambda_\nu}{d|\lambda_\nu|} \in \mathcal{K}$ $|\lambda_\nu|$-a.e., and moreover

$$\int_{\overline{\Omega}} j(\lambda_\nu, \epsilon) \geq \int_T \left(\int_{\overline{\Omega}} j(\lambda^t, \epsilon) \right) \nu(dt). \tag{3.51}$$

Proof. The idea is to show there exists a continuous function $F : V \to (W)^* \equiv W$ such that for every $\xi \in V$ we obtain a minorization $\langle F(\xi),\, \cdot \rangle \leq j^-(\,\cdot\,, \xi)$ on W where $j^- = -j$. Once this is established we define $h : \overline{\Omega} \times W \to \overline{\mathbb{R}}$ by

$$h(x, H) := j^-\big(H, \epsilon(x)\big) - \big\langle F\big(\epsilon(x)\big), H \big\rangle.$$

Since F is continuous and j^- is lower semi-continuous jointly on $W \times V$ by Proposition 3.6, we see by uniform continuity of ϵ that the function h is lower semi-continuous jointly on $\overline{\Omega} \times W$. Then, owing to non-negativity of h and its convexity together with positive one-homogeneity with respect to the second variable, the functional $\lambda \mapsto \int h\big(x, \lambda(dx)\big)$ is weakly-* lower semi-continuous on $\mathcal{M}\big(\overline{\Omega}; W\big)$ due to Reshetnyak's result, see Corollary 2.25. We observe that for any $\epsilon \in C(\overline{\Omega}; V)$

$$\int j^-\big(\lambda, \epsilon\big) = \int h\big(x, \lambda(dx)\big) + \int \big\langle F\big(\epsilon(x)\big), \lambda(dx) \big\rangle,$$

hence the functional $\lambda \mapsto \int j^-\big(\lambda, \epsilon\big)$ is a sum of a weakly-* lower semi-continuous functional on $\mathcal{M}\big(\overline{\Omega}; W\big)$ and a continuous linear functional (the function $F \circ \epsilon : \overline{\Omega} \to W$ is uniformly continuous), which through equality $\int j\big(\lambda, \epsilon\big) = -\int j^-\big(\lambda, \epsilon\big)$ furnishes the assertion (i). Similarly we can obtain inequality (3.51) by employing Corollary 2.27 for the functional $\lambda \mapsto \int h\big(x, \lambda(dx)\big)$. Since $\lambda^t \in \mathcal{M}(\overline{\Omega}; \mathcal{K}) \subset \mathcal{M}(\overline{\Omega}; W)$ for every $t \in T$ the right hand side of (3.51) is non-negative (see Remark 3.34) and so is the integral $\int_{\overline{\Omega}} j(\lambda_\nu, \epsilon)$, therefore $\lambda_\nu \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ by acknowledging Remark 3.34 again, which gives the assertion (ii).

To conclude the proof we must therefore show existence of the function F. First we show that for every $\xi \in V$ the proper, convex and lower semi-continuous function $j^-(\,\cdot\,, \xi) : W \to \overline{\mathbb{R}}$ is subdifferentiable at the origin, i.e. $\partial_1 j^-(0, \xi) \neq \varnothing$ where in this proof by $\partial_1 j^-$ we shall understand the subdifferential with respect to the first argument. By [140, Theorem 23.3] the scenario $\partial_1 j^-(0, \xi) = \varnothing$ can occur only if there exists a direction $\Delta H \in W$ such that the directional derivative with respect to the first argument $j^-_{\Delta H}(0, \xi)$

equals $-\infty$. Since $j^-(\,\cdot\,,\xi)$ is positively 1-homogeneous our argument for subdifferentiability at the origin amounts to verifying that $j^-(H,\xi) > -\infty$ for every H in a unit sphere in W. But this is trivial since we know that $j^-(\,\cdot\,,\xi)$ is proper for each ξ.

We have thus arrived at a multifunction $\Gamma : V \ni \xi \mapsto \partial_1 j^-(0,\xi) \in \big(2^W \backslash \varnothing\big)$ that is convex valued and closed valued. According to Michael's theorem (cf. Theorem 2.33) in order to show that there exists a continuous selection F of Γ it suffices to show that Γ is l.s.c (as a multifunction). According to Theorem 2.32 this boils down to proving that function $(\xi,\Delta H) \mapsto \chi^*_{\partial_1 j^-(0,\xi)}(\Delta H)$ is lower semi-continuous on $W \times V$. Theorem 23.2 in [140] says that $\chi^*_{\partial_1 j^-(0,\xi)}(\Delta H)$ is exactly the directional derivative of $j^-(\,\cdot\,,\xi)$ at $H = 0$ in the direction ΔH, but due to homogeneity of $j^-(\,\cdot\,,\xi)$ this derivative is precisely $j^-(\Delta H,\xi)$ and all boils down to showing lower semi-continuity of $j^- = -j$, which is guaranteed by Proposition 3.6 herein. □

For the next step we focus on the integral $\int j(\mathscr{H},\epsilon)\,d\mu$ as an energy potential, i.e. a functional in variable ϵ. For any $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\mathscr{H} \in L^\infty_\mu(\overline{\Omega};\mathcal{K})$ we define a functional on the Lebesgue space:

$$J_{\mu,\mathscr{H}}(\epsilon) := \int j(\mathscr{H},\epsilon)\,d\mu \qquad \text{for any } \epsilon \in L^p_\mu(\overline{\Omega};V). \tag{3.52}$$

The results below establish continuity of $J_{\mu,\mathscr{H}}$ and give an integral formula for the convex conjugate $J^*_{\mu,\mathscr{H}}$:

Proposition 3.36. *For a given Radon measure $\mu \in \mathcal{M}_+(\overline{\Omega})$ let $\mathscr{H} : \overline{\Omega} \to \mathcal{K}$ be a μ-measurable function. Then, $j\big(\mathscr{H}(\,\cdot\,),\,\cdot\,\big) : \overline{\Omega} \times V \to \mathbb{R}$ is a Carathéodory function, i.e.*

(i) for μ-a.e. $x \in \overline{\Omega}$ the function $j\big(\mathscr{H}(x),\,\cdot\,\big)$ is continuous;
(ii) for every $\xi \in V$ the function $j\big(\mathscr{H}(\,\cdot\,),\xi\big)$ is μ-measurable.

Proof. Statement (i) follows easily from assumption (A1) since every convex function that is finite on the whole finite dimensional space is automatically continuous. For any $\xi \in V$ the function $j\big(\mathscr{H}(\,\cdot\,),\xi\big)$ is a composition of upper semi-continuous function $j(\,\cdot\,,\xi)$ (cf. assumption (A3)) and μ-measurable function $\mathscr{H}$, hence the statement (ii). □

Corollary 3.37. *For any $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\mathscr{H} \in L^\infty_\mu(\overline{\Omega};\mathcal{K})$ the non-negative functional $J_{\mu,\mathscr{H}}$ is convex continuous on $L^p_\mu(\overline{\Omega};V)$. Its convex conjugate $J^*_{\mu,\mathscr{H}} : L^{p'}_\mu(\overline{\Omega};V) \to \overline{\mathbb{R}}$ reads*

$$J^*_{\mu,\mathscr{H}}(\sigma) = \int j^*(\mathscr{H},\sigma)\,d\mu \qquad \textit{for any } \sigma \in L^{p'}_\mu(\overline{\Omega};V).$$

Prior to the proof we make an observation:

Remark 3.38. It is worth to stress that one <u>cannot</u> write $J^*_{\mu,\mathscr{H}}(\sigma) = \int j^*(\mathscr{H}\mu,\sigma) = \int j^*(\lambda,\sigma)$ for $\lambda = \mathscr{H}\mu$ since $j^*(\,\cdot\,,\sigma)$ is not one-homogeneous, cf. Proposition 3.7. To put it differently, for a given $\sigma \in L^{p'}_\mu(\overline{\Omega};V)$ the integral $\int j^*(\mathscr{H},\sigma)\,d\mu$ depends not only on $\lambda = \mathscr{H}\mu$ but also on the particular choice of the positive measure μ, hence the symbol $J_{\mu,\mathscr{H}}$ rather than J_λ. An additional explanation behind this subtlety may be found in Remark 4.5.

Proof. Convexity of $J_{\mu,\mathscr{H}}$ follows directly from (A1). By one-homogeneity of $j(\,\cdot\,,\xi)$ and by Proposition 3.11 we have

$$j(\mathscr{H},\xi) \le \|c(\mathscr{H})\|_{L^\infty_\mu}\, j\Big(\frac{\mathscr{H}}{\|c(\mathscr{H})\|_{L^\infty_\mu}},\xi\Big) \le \|c(\mathscr{H})\|_{L^\infty_\mu}\,\frac{1}{p}\big(\rho(\xi)\big)^p \le \|c(\mathscr{H})\|_{L^\infty_\mu}\,\frac{(C_1)^p}{p}|\xi|^p$$

and thus $J_{\mu,\mathscr{H}}(\epsilon) \le \|c(\mathscr{H})\|_{L^\infty_\mu}\,\frac{(C_1)^p}{p}\big(\|\epsilon\|_{L^p_\mu}\big)^p$, which establishes continuity of $J_{\mu,\mathscr{H}}$, cf. Proposition 2.6.

The formula for $J^*_{\mu,\mathscr{H}}$ is due to the celebrated Rockafellar's result: any Carathéodory function is a normal convex integrand owing to [137, Lemma 2] and then assertion follows by [137, Theorem 2]. □

The subsequent result brings the local formula for maximizing the energy with respect to anisotropy: $\frac{1}{p}(\rho(\xi))^p = \max_{H\in\mathcal{K},\ c(H)\le 1} j(H,\xi)$ to the integral setting.

Proposition 3.39. *Let C_0 be a positive constant. For any $\epsilon \in C(\overline{\Omega};V)$ there holds*

$$\bar{J}(\epsilon) := \max\left\{\int j(\lambda,\epsilon)\ :\ \lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda)\le C_0\right\} = \frac{C_0}{p}\big(\|\rho(\epsilon)\|_{\infty,\Omega}\big)^p.$$

Proof. After representing λ as $\mathscr{H}\mu$ we obtain.

$$\bar{J}(\epsilon) = \sup_{\mu\in\mathcal{M}_+(\overline{\Omega}),\ \mathscr{H}\in L^\infty_\mu(\overline{\Omega};\mathcal{K})}\left\{\int j(\mathscr{H},\epsilon)\,d\mu\ :\ \int d\mu\ \le C_0,\ c(\mathscr{H})=1\ \mu\text{-a.e.}\right\}.$$

Let us fix a function ϵ. For any pair $\mathscr{H},\mu$ admissible above we find an estimate

$$\int j(\mathscr{H},\epsilon)\,d\mu \le \int \bar{j}(\epsilon)\,d\mu \le\ \|\bar{j}(\epsilon)\|_{\infty,\Omega}\int d\mu\ \le C_0\,\|\bar{j}(\epsilon)\|_{\infty,\Omega},$$

which yields $\bar{J}(\epsilon)\ \le\ C_0\,\|\bar{j}(\epsilon)\|_{\infty,\Omega}$. We shall show that the RHS of this inequality is attainable for a certain competitor $\bar{\lambda}_\epsilon$.

By Proposition 3.13 and continuity of ϵ the function $\bar{j}\big(\epsilon(\,\cdot\,)\big)$ is continuous on a compact set $\overline{\Omega}$ as well and thus there exists $\bar{x}\in\overline{\Omega}$ such that $\|\bar{j}(\epsilon)\|_{\infty,\Omega} = \bar{j}\big(\epsilon(\bar{x})\big)$. We put $\bar{\lambda}_\epsilon = C_0\,\bar{H}_{\epsilon(\bar{x})}\,\delta_{\bar{x}}$ where $\bar{H}_{\epsilon(\bar{x})}$ is any element in the non-empty set $\bar{\mathcal{K}}_1\big(\epsilon(\bar{x})\big)$ and $\delta_{\bar{x}}$ is the Dirac delta measure at $\bar{x}$. It is trivial to check that $\int c(\bar{\lambda}_\epsilon) = C_0$ whilst

$$\int j(\bar{\lambda}_\epsilon,\epsilon) = C_0\int j\big(\bar{H}_{\epsilon(\bar{x})},\epsilon(x)\big)\,\delta_{\bar{x}}(dx) = C_0\,j\big(\bar{H}_{\epsilon(\bar{x})},\epsilon(\bar{x})\big) = C_0\,\bar{j}\big(\epsilon(\bar{x})\big),$$

which proves that indeed $\bar{J}(\epsilon) = C_0\,\|\bar{j}(\epsilon)\|_{\infty,\Omega}$ or that $\bar{J}(\epsilon) = \frac{C_0}{p}\,\big(\|\rho(\epsilon)\|_{\infty,\Omega}\big)^p$. □

The dual counterpart of Proposition 3.39, i.e. translating the relation $\frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = \inf_{H\in\mathcal{K},\ c(H)\le 1} j^*(H,\sigma)$ to the integral setting is far more subtle. First of all, we must work with the decomposition $\lambda = \mathscr{H}\mu$ as in (3.49), see Remark 3.38. Then, for fixed μ and $\sigma:\overline{\Omega}\to V$, the objective is to minimize $\int j^*(\mathscr{H},\sigma)\,d\mu$ with respect to the normalized anisotropy function $\mathscr{H}\in L^\infty_\mu(\overline{\Omega};\mathcal{K}_1)$.

Proposition 3.40. *For a given Radon measure $\mu\in\mathcal{M}_+(\overline{\Omega})$ let $\sigma:\overline{\Omega}\to V$ be a μ-measurable function. Then*

$$\inf\left\{\int j^*(\mathscr{H},\sigma)\,d\mu\ :\ \mathscr{H}\in L^\infty_\mu(\overline{\Omega};\mathcal{K}),\ c(\mathscr{H})\le 1\ \ \mu\text{-a.e.}\right\} = \frac{1}{p'}\int\big(\rho^0(\sigma)\big)^{p'}d\mu. \quad (3.53)$$

Moreover, if $\int \big(\rho^0(\sigma)\big)^{p'} d\mu < \infty$ *the minimum above is achieved and the set of solutions is precisely:*

$$\Big\{ \mathscr{H} : \overline{\Omega} \to \mathcal{K}_1 \ : \ \mathscr{H} \text{ is } \mu\text{-measurable}, \ \ \mathscr{H}(x) \in \underline{\mathcal{K}}_1\big(\sigma(x)\big) \ \ \text{for } \mu\text{-a.e. } x \Big\}. \tag{3.54}$$

Proof. Considering the relation $\frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = \inf_{H\in\mathcal{K}_1} j^*(H,\sigma)$ and definiton of $\mathcal{K}_1$ the inequality $\geq$ is straightforward. The same goes for the equality if $\int(\rho^0(\sigma))^{p'}d\mu = \infty$. Otherwise, we have $\rho^0(\sigma) < \infty$ μ-a.e. and then, thanks to definition of the set $\underline{\mathcal{K}}_1(\,\cdot\,)$, the whole claim is proved if we show that the set (3.54) is non-empty. This part is dealt with in Lemma 3.41 below. □

The most challenging part of proving Proposition 3.40 is showing non-emptiness of (3.54), namely that point-wise the anisotropy $\mathscr{H}(x)$ may be chosen optimally for $\sigma(x)$ in a μ-measurable fashion:

Lemma 3.41. *For a given Radon measure* $\mu \in \mathcal{M}_+(\overline{\Omega})$ *let* $\sigma : \overline{\Omega} \to V$ *be a* μ*-measurable function. We consider a compact-valued and convex-valued multifunction* $\Gamma_\sigma : \overline{\Omega} \to 2^{\mathcal{K}_1}\backslash\varnothing$ *as below*

$$\Gamma_\sigma(x) := \underline{\mathcal{K}}_1\big(\sigma(x)\big) = \Big\{ H \in \mathcal{K}_1 : \bar{j}^*\big(\sigma(x)\big) = j^*\big(H,\sigma(x)\big) \Big\}.$$

Then there exists a μ*-measurable selection* $\mathscr{H}_\sigma : \overline{\Omega} \to \mathcal{K}_1$ *of the multifunction* Γ_σ*, namely*

$$\mathscr{H}_\sigma(x) \in \underline{\mathcal{K}}_1\big(\sigma(x)\big) \qquad \text{for } \mu\text{-a.e. } x \in \overline{\Omega}.$$

Proof. Let $\mathscr{B}(W)$ stand for the Borel σ-algebra on W, while $\mathscr{B}_\mu(\overline{\Omega})$ is the μ-complete σ-algebra generated by the Borel sets and μ-negligible sets, i.e. μ-measurability of σ is meant precisely with respect to $\mathscr{B}_\mu(\overline{\Omega})$, see Section 2.2.

We shall begin by showing that the function $j^*\big(\,\cdot\,,\sigma(\,\cdot\,)\big) : W \times \overline{\Omega} \to \overline{\mathbb{R}}$ is measurable with respect to the product σ-algebra $\mathscr{B}(W) \otimes \mathscr{B}_\mu(\overline{\Omega})$. The function $j^*\big(\,\cdot\,,\sigma(\,\cdot\,)\big)$ may be written as a composition map $j^* \circ g$ where $g = (\mathrm{id},\sigma) : W \times \overline{\Omega} \to W \times V$, where id is the identity map. Owing to measurability of σ it is straightforward to check that g is measurable with respect to $\mathscr{B}(W) \otimes \mathscr{B}_\mu(\overline{\Omega})$. Then, the function $j^*\big(\,\cdot\,,\sigma(\,\cdot\,)\big)$ is measurable with respect to $\mathscr{B}(W) \otimes \mathscr{B}_\mu(\overline{\Omega})$ as a composition of a lower semi-continuous function and a measurable function. The proof that the mapping $(W \times \overline{\Omega}) \ni (H,x) \mapsto \bar{j}^*\big(\sigma(x)\big)$ is $\mathscr{B}(W) \otimes \mathscr{B}_\mu(\overline{\Omega})$-measurable is identical since $\bar{j}^*$ is again lower semi-continuous.

Owing to equality (3.14) the graph $\mathcal{G}$ of multifunction Γ_σ reads

$$\mathcal{G} = \Big\{ (x,H) \in \overline{\Omega} \times \mathcal{K}_1 \ : \ j^*\big(H,\sigma(x)\big) \leq \bar{j}^*\big(\sigma(x)\big) \Big\}.$$

We introduce a subset of V:

$$V_{\mathrm{fin}} := \Big\{ \sigma \in V \ : \ \bar{j}^*(\sigma) < \infty \Big\} = \bigcup_{n=1}^{\infty} \Big\{ \sigma \in V \ : \ \bar{j}^*(\sigma) \leq n \Big\}$$

which is a F_σ set and thus Borel measurable; consequently the sets: $\overline{\Omega}_{\mathrm{fin}} = \sigma^{-1}(V_{\mathrm{fin}})$, $\overline{\Omega}_\infty = \sigma^{-1}(V \setminus V_{\mathrm{fin}})$ are a μ-measurable partitioning of $\overline{\Omega}$. The graph $\mathcal{G}$ may be thus split into two parts as follows:

$$\mathcal{G} = \mathcal{G}_{\mathrm{fin}} \cup \mathcal{G}_\infty, \ \ \mathcal{G}_{\mathrm{fin}} = \Big\{ (x,H) \in \overline{\Omega}_{\mathrm{fin}} \times \mathcal{K}_1 \ : \ j^*\big(H,\sigma(x)\big) - \bar{j}^*\big(\sigma(x)\big) \leq 0 \Big\}, \ \ \mathcal{G}_\infty = \overline{\Omega}_\infty \times \mathcal{K}_1.$$

Both the sets $\mathcal{G}_{\text{fin}}$ and $\mathcal{G}_\infty$ are measurable with respect to $\mathscr{B}_\mu(\overline{\Omega}) \otimes \mathscr{B}(W)$. It is clear for $\mathcal{G}_\infty$, whilst for the $\mathcal{G}_{\text{fin}}$ it follows from the fact that $j^*\big(\,\cdot\,, \sigma(\,\cdot\,)\big) - \bar{j}^{\,*}\big(\sigma(\,\cdot\,)\big)$ is a difference of two $\mathscr{B}_\mu(\overline{\Omega}) \otimes \mathscr{B}(W)$-measurable functions as it was proved above.

Measurability of the graph $\mathcal{G}$ with respect to the product σ-algebra $\mathscr{B}_\mu(\overline{\Omega}) \otimes \mathscr{B}(W)$ follows. Since $\mathscr{B}_\mu(\overline{\Omega})$ is μ-complete from Theorem 2.38 we infer that Γ is a measurable multifunction with respect to $\mathscr{B}_\mu(\overline{\Omega})$. Our assertion follows from Theorem 2.36. □

The proof of the lemma simplifies greatly once the condition $(\text{A}_{\text{ellip}})$ is assumed, in particular it does not necessitate the advanced result in Theorem 2.38. The present author deems it worthwhile to display:

Proof under the additional assumption $(\text{A}_{\text{ellip}})$. It suffices to prove that the multifunction $V \ni \tilde{\sigma} \mapsto \underline{\mathcal{K}}_1(\tilde{\sigma}) \subset \mathcal{K}_1$ is upper semi-continuous. Then it is also Borel measurable according to Theorem 2.37 and thus, owing to Theorem 2.36 there exists a Borel measurable selection $\underline{H} : V \to \mathcal{K}_1$, i.e. $\underline{H}(\tilde{\sigma}) \in \underline{\mathcal{K}}_1(\tilde{\sigma})$ for every $\tilde{\sigma} \in V$. Then $\underline{\mathscr{H}}_\sigma := \underline{H} \circ \sigma : \overline{\Omega} \to \mathcal{K}_1$ is μ-measurable as a composition of a Borel measurable and a μ-measurable function.

We now show upper semi-continuity of $\tilde{\sigma} \mapsto \underline{\mathcal{K}}_1(\tilde{\sigma})$. Since $\mathcal{K}_1$ is compact, by virtue of Theorem 2.31 it is enough to show that the graph $\mathcal{G}$ of multifunction Γ_σ is closed, while thanks to equality (3.14) one can write

$$\mathcal{G} = \Big\{(\tilde{\sigma}, H) \in V \times \mathcal{K}_1 \,:\, j^*(H, \tilde{\sigma}) - \bar{j}^{\,*}(\tilde{\sigma}) \leq 0\Big\}.$$

Owing to condition $(\text{A}_{\text{ellip}})$ the function $\bar{j}^{\,*} = \frac{1}{p'}(\rho^0(\,\cdot\,))^{p'}$ is real-valued and continuous, see Proposition 3.13. The function $j^* : \mathcal{K}_1 \times V \to \overline{\mathbb{R}}$ is jointly lower semi-continuous therefore so is the mapping $(H, \tilde{\sigma}) \mapsto j^*(H, \tilde{\sigma}) - \bar{j}^{\,*}(\tilde{\sigma})$. As a result the graph $\mathcal{G}$ is closed in $V \times \mathcal{K}_1$ which finishes the proof. □

Chapter 4

The family of Free Material Design problems

The main purpose of this chapter is to establish the mathematical theory of Free Material Design (FMD) – the optimal design problem posed for the first time in [20, 136] and recently revisited and developed in [64, 65, 62, 69, 66]. Following the footsteps of [37] in the first Sections 4.1 - 4.4 we will be working with an abstract universal framework of (FMD) which shall be then specified in Section 4.5 for: #1 classical 2D and 3D elasticity, #2 bending of thin Kirchhoff plates, #3 stationary heat conduction.

The methodology employed can be summarized as generalization of the methods of Mass Optimization Problem [35, 37, 42] by means of the anisotropy apparatus developed in Chapter 3 of this thesis. As a consequence, in this chapter (and only in this chapter) the reader will find variations of numerous results and arguments already established in the works [35, 37, 42]. For the thesis to be self-contained as much as possible all the proofs will be given, even if the result is an unaltered repetition of the original version (suitable references shall be made).

4.1. Formulation of the Free Material Design problem

Let us assume a *design domain* Ω being a bounded open connected set in $\mathbb{R}^d$. For the *Dirichlet zone* Υ_0 we take any compact subset of $\overline{\Omega}$. Let U and V be any finite dimensional spaces endowed with scalar products, both denoted by $\langle \cdot\,, \cdot \rangle$. We choose a linear continuous operator

$$A : \mathcal{D}\big(\mathbb{R}^d \setminus \Upsilon_0; U\big) \to C_0\big(\mathbb{R}^d \setminus \Upsilon_0; V\big).$$

As a *load* (or a heat source) we shall understand a U-valued distribution $\mathscr{F} \in \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U)$ that is supported on $\overline{\Omega}$, namely $\operatorname{spt} \mathscr{F} \subset \overline{\Omega}$. Finally, for another finite dimensional space W, we assume a closed convex cone $\mathcal{K} \subset W$, an energy function $j : \mathcal{K} \times V \to \mathbb{R}_+$ satisfying conditions (A1) - (A4) formulated in Section 3.1 and a cost function $c : \mathcal{K} \to \mathbb{R}_+$ being a restriction of a norm on W.

An elastic body (or a heat conductor) will be identified with the field $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ which by decomposition $\lambda = \mathscr{H}\mu$ conveys information on the shape/mass μ and anisotropy $\mathscr{H}$. By *compliance* of the body λ subject to the load $\mathscr{F}$ we understand the minus potential energy stored in the system:

$$\mathcal{C} = \mathcal{C}(\lambda) := \sup \left\{ \mathscr{F}(u) - \int j\big(\lambda, Au\big) \;:\; u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U) \right\}, \tag{4.1}$$

where function u plays a role of a state variable: the displacement field of an elastic body or the temperature field in a conductor.

The problem of minimizing the compliance by finding an optimal constitutive field λ of total cost not exceeding $C_0 > 0$ shall be referred to as the *Free Material Design* problem:

$$\mathcal{C}_{\min} = \inf\left\{\mathcal{C}(\lambda) \;:\; \lambda \in \mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda) \le C_0\right\}. \tag{FMD}$$

Thanks to the analysis of the functional $\lambda \mapsto \int j(\lambda,\epsilon)$ carried out in Section 3.4 we immediately obtain an existence result:

Proposition 4.1. *Whenever $\mathcal{C}_{\min} < \infty$ there exists a solution $\check{\lambda}$ of the* (FMD) *problem.*

Proof. Compliance functional $\mathcal{C}$ is a point-wise supremum of functionals $\lambda \mapsto \mathscr{F}(u) - \int j(\lambda, Au)$ that are convex and weakly-* lower semi-continuous on $\mathcal{M}\big(\overline{\Omega};W\big)$ by virtue of Theorem 3.35. Therefore $\mathcal{C}$ is itself convex and weakly-* lower semi-continuous, cf. Proposition 2.8. Since c is a norm, minimization in (FMD) runs over a non-empty compact subset of $\mathcal{M}\big(\overline{\Omega};W\big)$ (see Remark 3.34) and the assertion follows by the Direct Method of Calculus of Variations, cf. Theorem 2.3. □

One can observe that (FMD) problem thus posed is subject to very mild assumptions, e.g. concerning the operator A as well as the load $\mathscr{F}$. This is reflected in the prerequisite of Proposition 4.1: it may happen that $\mathcal{C}(\lambda) = \infty$ for every feasible λ. Nonetheless, we shall carry on with this general framework throughout the next subsection to extract some versatile results that will be vital in Chapter 6. Afterwards, in Section 4.3 we give a set of assumptions under which $\mathcal{C}_{\min} < \infty$ and solution of (FMD) does exists.

4.2. The passage to a pair of convex problems $(\mathcal{P})$ and $(\mathcal{P}^*)$. Constructing solution of the (FMD) problem

In this section we examine a connection of (FMD) problem to a pair of convex problems:

$$Z := \sup\left\{\mathscr{F}(u) \;:\; u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U),\ \ \rho(Au) \le 1 \ \text{ in } \overline{\Omega}\right\} \tag{$\mathcal{P}$}$$

and, what shall appear to be the dual of $(\mathcal{P})$,

$$\inf\left\{\int \rho^0(\varsigma) \;:\; \varsigma \in \mathcal{M}\big(\overline{\Omega};V\big),\ \ A^*\varsigma = \mathscr{F} \ \text{ in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U)\right\} \tag{$\mathcal{P}^*$}$$

where the Borel measure $\rho^0(\varsigma)$ is understood as in (2.9). The adjoint A^* is computed for A treated as an operator from $\mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U)$ to $C(\overline{\Omega};\mathbb{R}^d)$ in which the space $C_0(\mathbb{R}^d \setminus \Upsilon_0; V)$ is continuously embedded (in the sense of restrictions to $\overline{\Omega}$), i.e. we consider a linear continuous operator $A^* : \mathcal{M}(\overline{\Omega};V) \to \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U)$.

Above $\rho : V \to \mathbb{R}$ and $\rho^0 : V \to \overline{\mathbb{R}}$ are the mutually polar closed gauge functions studied in Chapter 3. They satisfy the following formulas:

$$\frac{1}{p}\big(\rho(\xi)\big)^p = \max_{H\in\mathcal{K},\ c(H)\le 1} j(H,\xi), \qquad \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = \inf_{H\in\mathcal{K},\ c(H)\le 1} j^*(H,\sigma); \tag{4.2}$$

let us recall that without the ellipticity assumption ($\mathrm{A}_{\mathrm{ellip}}$) function ρ^0 admits infinite values.

We present the first connection between problems ($\mathcal{P}$) and (FMD), which may be found as generalization of [37, Proposition 2] towards anisotropy design:

Lemma 4.2. *The minimum value of compliance in* (FMD) *equals*

$$\mathcal{C}_{\min} = \frac{1}{p'\, {C_0}^{p'-1}}\ Z^{\,p'} \tag{4.3}$$

where $Z = \sup \mathcal{P}$, although both Z and $\mathcal{C}_{\min}$ can equal ∞. Moreover, with $\check{\lambda}$ being any solution of (FMD) *the problem* (4.1) *for $\lambda = \check{\lambda}$ admits a maximizing sequence u_n that satisfies the uniform estimate:*

$$\|\rho(Au_n)\|_{\infty,\Omega} \leq \left(\frac{Z}{C_0}\right)^{p'-1}. \tag{4.4}$$

Proof. By plugging definition (4.1) into (FMD) problem we arrive at a min-max problem:

$$\begin{aligned}\mathcal{C}_{\min} &= \inf_{\lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda)\leq C_0}\ \sup_{u\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)} \left\{\mathscr{F}(u) - \int j\big(\lambda, Au\big)\right\}\\ &= \sup_{u\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)}\ \min_{\lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda)\leq C_0} \left\{\mathscr{F}(u) - \int j\big(\lambda, Au\big)\right\} = \sup_{u\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)} \left\{\mathscr{F}(u) - \bar{J}(Au)\right\}\end{aligned}$$

where functional $\bar{J} : C(\overline{\Omega};V) \to \mathbb{R}_+$ is defined in Proposition 3.39. Above we have interchanged the order of inf and sup by virtue of Ky Fan's result, see Theorem 2.5: convexity and weak-* lower semi-continuity of $\lambda \mapsto \mathscr{F}(u) - j\big(\lambda, Au\big)$ follows directly from Proposition 3.35 and compactness of the set $\big\{\lambda \in \mathcal{M}(\overline{\Omega};\mathcal{K}) : \int c(\lambda) \leq C_0\big\}$ follows from Proposition 3.33 and the fact that c is a norm; concavity of the functional $u \mapsto \mathscr{F}(u) - j\big(\lambda, Au\big)$ is a consequence of linearity of A and assumption (A1) on j.

By virtue of Proposition 3.39 there holds $\bar{J}(Au) = \frac{C_0}{p}\left(\|\rho(Au)\|_{\infty,\Omega}\right)^p$. As the next step we use a technique that was applied in e.g. [35]: by substitution $u = t\,u_1$ we obtain

$$\begin{aligned}\mathcal{C}_{\min} &= \sup_{u\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)} \left\{\mathscr{F}(u) - \frac{C_0}{p}\left(\|\rho(Au)\|_{\infty,\Omega}\right)^p\right\}\\ &= \sup_{t\geq 0}\ \sup_{u_1\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)} \left\{\big(\mathscr{F}(u_1)\big)\,t - \frac{C_0}{p}\,t^p :\ \|\rho(Au_1)\|_{\infty,\Omega} \leq 1\right\}\\ &= \sup_{t\geq 0}\left\{Z\,t - \frac{C_0}{p}\,t^p\right\} = \frac{1}{p'\,{C_0}^{p'-1}}\ Z^{\,p'}\end{aligned}$$

where the maximum with respect to t was found for $\bar{t} = \big(Z/C_0\big)^{p'-1}$. Based on the above and (4.2) we may write

$$\begin{aligned}\mathcal{C}_{\min} = \mathcal{C}(\check{\lambda}) &= \sup_{u\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)} \left\{\mathscr{F}(u) - \int j\big(\check{\lambda}, Au\big)\right\}\\ &\geq \sup_{u\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)} \left\{\mathscr{F}(u) - \int j\big(\check{\lambda}, Au\big)\ :\ \|\rho(Au)\|_{\infty,\Omega} \leq \bar{t}\right\}\\ &\geq \sup_{u\in\mathcal{D}(\mathbb{R}^d\setminus\Upsilon_0;U)} \left\{\mathscr{F}(u) - \frac{C_0}{p}\left(\|\rho(Au)\|_{\infty,\Omega}\right)^p\ :\ \|\rho(Au)\|_{\infty,\Omega} \leq \bar{t}\right\} = \mathcal{C}_{\min}.\end{aligned}$$

The second inequality once again utilizes Proposition 3.39: $\check{\lambda}$ satisfies $\int c(\check{\lambda}) \leq C_0$. Ultimately we have equalities everywhere and the "moreover part" follows. □

Next, through classical tools of convex analysis we establish duality between $(\mathcal{P})$ and $(\mathcal{P}^*)$:

Proposition 4.3. *There holds a zero-gap equality*

$$Z = \sup \mathcal{P} = \inf \mathcal{P}^*$$

whereas the minimum is achieved whenever $Z < \infty$.

Proof. The assertion follows by a direct application of Theorem 2.12. Indeed, let us take:

$$X = \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U), \qquad Y = C(\overline{\Omega}; V), \qquad \Lambda = A$$

and

$$\Phi(u) = -\mathscr{F}(u), \qquad \Psi(\epsilon) = \chi_{\mathscr{C}}(\epsilon)$$

where $\mathscr{C} = \big\{\epsilon \in C(\overline{\Omega}; V) \,:\, \epsilon(x) \in \mathrm{C} \quad \forall\, x \in \overline{\Omega}\big\}$ for $\mathrm{C} = \big\{\xi \in V : \rho(\xi) \leq 1\big\}$. We readily recognize the maximization problem on the LHS of (2.2) as $(\mathcal{P})$. For any $\varsigma \in Y^* = \mathcal{M}(\overline{\Omega}; V)$ it is straightforward to check that

$$\Phi^*(-\Lambda^*\varsigma) = \begin{cases} 0 & \text{if} \quad A^*\varsigma = \mathscr{F} \ \text{ in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U) \\ \infty & \text{otherwise} \end{cases}$$

while, by acknowledging that $\chi^*_{\mathrm{C}}(\sigma) = \rho^0(\sigma)$ (see (2.3)), equality $\int \rho^0(\varsigma) = \Psi^*(\varsigma)$ follows directly from Corollary 2.24. The minimization problem on the RHS of (2.2) is thus precisely $(\mathcal{P}^*)$. The set C contains a neighbourhood $\mathscr{W}$ of the origin in V (cf. the "moreover part" in Proposition 3.11) thus for $u_0 = 0 \in C(\overline{\Omega}; V)$ clearly $\Phi(u_0) < \infty$ and Φ is continuous at $\Lambda u_0 = 0$ in the topology of uniform convergence: $\chi_{\mathscr{C}}(\epsilon)$ is zero for any ϵ such that $\epsilon(x) \in \mathscr{W}$ for every $x \in \overline{\Omega}$. □

The remaining goal of this subsection is to give a recipe for construction of $\check{\lambda}$ optimal for the (FMD) problem by exploiting solution $\hat{\varsigma}$ of $(\mathcal{P}^*)$. Beforehand, we require a dual definition of compliance $\mathcal{C}(\lambda)$ that will unlock a more direct link between problems (FMD) and $(\mathcal{P}^*)$:

Proposition 4.4. *For any representation of $\lambda = \mathscr{H}\mu$ of $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ such that $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\mathscr{H} \in L^\infty_\mu(\overline{\Omega}; \mathcal{K})$ there holds*

$$\mathcal{C}(\lambda) = \inf \left\{ \int j^*\big(\mathscr{H}, \sigma\big)\, d\mu \;:\; \sigma \in L^{p'}_\mu(\overline{\Omega}; V), \quad A^*(\sigma\mu) = \mathscr{F} \ \text{ in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U) \right\} \tag{4.5}$$

whereas the minimum is achieved whenever $\mathcal{C}(\lambda) < \infty$.

Remark 4.5. Should it exist, solution σ of the problem (4.5) depends on the particular choice of μ and $\mathscr{H}$ that gives $\lambda = \sigma\mu$, cf. Remark 3.38. For instance, if for given μ we put $\mu_1 = \alpha\mu$, $\mathscr{H}_1 = \frac{1}{\alpha}\mathscr{H}$ for some $\alpha > 0$, solution σ_1 for this pair would satisfy the scaling property: $\sigma_1 = \frac{1}{\alpha}\sigma$. This can be inferred by combining the form of constraint $A^*(\sigma\mu) = \mathscr{F}$ with property (3.7) of $j^*(\,\cdot\,, \sigma)$. To put it differently it is the field $\varsigma = \sigma\mu$ that is invariant

of the chosen representation $\lambda = \mathscr{H}\mu$. In the context of elasticity ς may be considered an *absolute stress field* whilst $\sigma = \frac{d\varsigma}{d\mu}$ (the Radon-Nikodym derivative) can be interpreted as the *relative stress field*, i.e. relative with respect to the elastic material's distribution μ.

Proof. Once more we shall utilize Theorem 2.12, this time to link the supremum in the definition (4.1) of $\mathcal{C}(\lambda)$ and the minimum in the assertion. We take X, Φ as in the proof of Proposition 4.3, while $Y = L^p_\mu(\overline{\Omega}; V)$ and $\Psi(\epsilon) = J_{\mu,\mathscr{H}}(\epsilon)$, see (3.52). We again put $\Lambda u := Au$ although now Λ is a linear operator from $\mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U)$ to $Y = L^p_\mu(\overline{\Omega}; V)$ which affects the adjoint: $\Lambda^*\sigma = A^*(\sigma\mu)$ for $\sigma \in L^{p'}_\mu(\overline{\Omega}; V)$ where A^* was computed for the operator $A : \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U) \to C(\overline{\Omega}; V)$. Owing to Corollary 3.37 we deduce continuity of Ψ (on the whole Y) as well as the formula $\Psi^*(\sigma) = \int j^*\big(\mathscr{H}, \sigma\big)\, d\mu$. The rest of the proof is analogous to the proof of Proposition 4.3. □

Before stating the main theorem of this section let us recall the definition of the set of normalized anisotropy parameters that are optimally designed for $\sigma \in V$:

$$\underline{\mathcal{K}}_1(\sigma) = \Big\{ H \in \mathcal{K} \,:\, c(H) \le 1,\ j^*(H, \sigma) = \bar{j}^*(\sigma) \Big\}$$

while $\bar{j}^*(\sigma) = \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'}$. Thus, in the case when $\bar{j}^*(\sigma) < \infty$ (see Definition 2.1), the convex compact set $\underline{\mathcal{K}}_1(\sigma)$ consists of solutions of the minimization problem in (4.2), while $\underline{\mathcal{K}}_1(\sigma) = \big\{ H \in \mathcal{K} \,:\, c(H) \le 1 \big\}$ otherwise.

Theorem 4.6. *Assuming that $0 < \mathcal{C}_{\min} < \infty$ let us consider a triple $(\hat{\mu}, \hat{\sigma}, \hat{\mathscr{H}})$ such that $\hat{\mu} \in \mathcal{M}_+(\overline{\Omega})$, $\hat{\sigma} \in L^1_{\hat{\mu}}(\overline{\Omega}; V)$ and $\hat{\mathscr{H}} \in L^\infty_{\hat{\mu}}(\overline{\Omega}; \mathcal{K})$. We define a triple $(\check{\mu}, \check{\sigma}, \check{\mathscr{H}})$ as below:*

$$\check{\mu} = \frac{C_0}{Z}\,\hat{\mu}, \qquad \check{\sigma} = \frac{Z}{C_0}\,\hat{\sigma}, \qquad \check{\mathscr{H}} = \hat{\mathscr{H}}. \tag{4.6}$$

Then, the two conditions are equivalent:

- *(i) the measure $\hat{\varsigma} = \hat{\sigma}\hat{\mu}$ solves problem ($\mathcal{P}^*$) with $\rho^0(\hat{\sigma}) = 1$ $\hat{\mu}$-a.e. whilst $\hat{\mathscr{H}}(x) \in \underline{\mathcal{K}}_1\big(\hat{\sigma}(x)\big)$ for μ-a.e. x;*
- *(ii) the measure $\check{\lambda} = \check{\mathscr{H}}\check{\mu}$ solves the* (FMD) *problem with $c(\check{\mathscr{H}}) = 1$ $\check{\mu}$-a.e. whilst $\check{\sigma}$ solves the minimization problem in* (4.5) *for $\mu = \check{\mu}$ and $\mathscr{H} = \check{\mathscr{H}}$.*

Proof. Let us first assume that the triple $(\hat{\mu}, \hat{\sigma}, \hat{\mathscr{H}})$ satisfies (i). Since $\rho^0(\hat{\sigma}) = 1$ $\hat{\mu}$-a.e. it is straightforward that $c(\hat{\mathscr{H}}) = 1$ $\hat{\mu}$-a.e. as well and the same concerns $\check{\mathscr{H}}$. We verify that $\check{\lambda} = \check{\mathscr{H}}\check{\mu}$ is a feasible constitutive field by computing the total cost:

$$\int c(\check{\lambda}) = \int c(\check{\mathscr{H}})\, d\check{\mu} = \int d\check{\mu} = \frac{C_0}{Z}\int d\hat{\mu} = \frac{C_0}{Z}\int \rho^0(\hat{\varsigma}) = \frac{C_0}{Z}\big(\min \mathcal{P}^*\big) = C_0.$$

In order to prove that $\check{\lambda}$ is a solution for (FMD) it suffices to show that $\mathcal{C}(\check{\lambda}) \le \mathcal{C}_{\min}$ where $\mathcal{C}_{\min} = Z^{p'}/\big(p'\, {C_0}^{p'-1}\big)$ by Lemma 4.2. We observe that $\hat{\mu}$-a.e. $\rho^0(\check{\sigma}) = \frac{Z}{C_0}\rho^0(\hat{\sigma}) = \frac{Z}{C_0}$. Since there holds $\check{\sigma}\check{\mu} = \hat{\sigma}\hat{\mu} = \hat{\varsigma}$ the distributional equation $A^*(\check{\sigma}\check{\mu}) = \mathscr{F}$ is satisfied owing to admissibility of $\hat{\sigma}\hat{\mu}$ for ($\mathcal{P}^*$). Due to the p'-homogeneity of j^* with respect to the second argument (cf. Proposition 3.7) the field $\check{\mathscr{H}} = \hat{\mathscr{H}}$ is a measurable selection for both

$x \mapsto \underline{\mathcal{K}}_1\big(\hat{\sigma}(x)\big)$ and $x \mapsto \underline{\mathcal{K}}_1\big(\check{\sigma}(x)\big)$. Then, by the dual definition of compliance (4.5) we may estimate

$$\mathcal{C}(\check{\lambda}) \leq \int j^*\big(\check{\mathscr{H}}, \check{\sigma}\big) d\check{\mu} = \int \frac{1}{p'}\Big(\rho^0\big(\check{\sigma}\big)\Big)^{p'} d\check{\mu} = \int \frac{1}{p'}\left(\frac{Z}{C_0}\right)^{p'} d\check{\mu} = \mathcal{C}_{\min},$$

where in the first equality we have used the fact that $\check{\mathscr{H}}(x) \in \underline{\mathcal{K}}_1\big(\check{\sigma}(x)\big)$ for $\hat{\mu}$-a.e. x, see Proposition 3.40; the last equality is due to $\int d\check{\mu} = C_0$. This proves minimality of $\check{\lambda}$ and we have only equalities in the chain above, which shows that $\check{\sigma}$ solves the minimization problem (4.5) for $\mu = \check{\mu}$ and $\mathscr{H} = \check{\mathscr{H}}$.

Conversely we assume that the triple $(\check{\mu}, \check{\sigma}, \check{\mathscr{H}})$ is a solution of the (FMD) problem while $c(\check{\mathscr{H}}) = 1 \ \ \check{\mu}$-a.e. The Hölder inequality furnishes

$$\int \rho^0(\check{\sigma})\, d\check{\mu} \leq \left(\int d\check{\mu}\right)^{1/p}\left(\int \big(\rho^0(\check{\sigma})\big)^{p'} d\check{\mu}\right)^{1/p'} \leq {C_0}^{1/p}\left(\int \big(\rho^0(\check{\sigma})\big)^{p'} d\check{\mu}\right)^{1/p'} \tag{4.7}$$

and the inequalities are equalities if and only if $\rho^0(\check{\sigma})$ is $\check{\mu}$-a.e. and $\int d\check{\mu} = C_0$ (note that $\check{\sigma}$ cannot be the zero function since $\mathcal{C}_{\min} > 0$). Based on the fact that $\check{\lambda} = \check{\mathscr{H}}\check{\mu}$ is a solution of (FMD) and $\check{\sigma}$ is a minimizer in (4.5) we may estimate:

$$\begin{aligned}\mathcal{C}_{\min} = \mathcal{C}(\check{\lambda}) = \int j^*(\check{\mathscr{H}}, \check{\sigma})\, d\check{\mu} &\geq \int \frac{1}{p'}\big(\rho^0(\check{\sigma})\big)^{p'}\, d\check{\mu} \\ &\geq \frac{1}{p' {C_0}^{p'/p}}\left(\int \rho^0(\check{\sigma})\, d\check{\mu}\right)^{p'} \geq \frac{Z^{p'}}{p' {C_0}^{p'/p}} = \mathcal{C}_{\min},\end{aligned}$$

where in the last inequality we use the fact that $\check{\varsigma} = \check{\sigma}\check{\mu}$ is admissible for $(\mathcal{P}^*)$. We see that above we have equalities everywhere, which implies several facts. First, we have $Z = \int \rho^0(\check{\varsigma})$, which shows that $\check{\varsigma} = \check{\sigma}\check{\mu} = \hat{\sigma}\hat{\mu} = \hat{\varsigma}$ solve $(\mathcal{P}^*)$. Then, by Hölder inequality (4.7) and the comment below it, we obtain that $\int d\check{\mu} = C_0$ and $\rho^0(\check{\sigma}) = t = \text{const}$ $\check{\mu}$-a.e. Combining those two facts we obtain $\rho^0(\check{\sigma}) = \frac{Z}{C_0}$ and consequently $\rho^0(\hat{\sigma}) = 1$. As the last information from the chain of equalities we make use of Proposition 3.40 implying that $\check{\mathscr{H}}(x) \in \underline{\mathcal{K}}_1\big(\check{\sigma}(x)\big) = \underline{\mathcal{K}}_1\big(\hat{\sigma}(x)\big)$ for $\hat{\mu}$-a.e. x and thus the same holds for $\check{\mathscr{H}}$. The proof is complete. □

The method of constructing solutions to (FMD) is already at hand owing to Lemma 3.41 on μ-measurable minimizers $\mathscr{H}$:

Corollary 4.7. *Assuming that $Z < \infty$, for $\varsigma \in \mathcal{M}(\overline{\Omega}; V)$ being any solution of $(\mathcal{P}^*)$ the measure $\mu = \rho^0(\varsigma)$ is Radon and we may define $\sigma = \frac{d\varsigma}{d\mu} \in L^\infty_\mu(\overline{\Omega}; V)$.*

Let $\mathscr{H} \in L^\infty_\mu(\overline{\Omega}; \mathcal{K})$ be chosen as any μ-measurable selection of the multifunction $\overline{\Omega} \ni x \mapsto \underline{\mathcal{K}}_1\big(\sigma(x)\big)$, which always exists. Then, the constitutive field

$$\lambda = \frac{C_0}{Z}\, \mathscr{H}\mu$$

solves the (FMD) *problem.*

Proof. The Borel measure $\rho^0(\varsigma)$ (cf. (2.9)) is finite since $\int \rho^0(\varsigma) = Z < \infty$ therefore μ is a Radon measure and $\sigma = \frac{d\varsigma}{d\mu}$ is a well-defined μ-measurable function. It is then straightforward to check that $\rho^0(\sigma) = 1 \ \ \mu$-a.e. and in particular $\sigma \in L^\infty_\mu(\overline{\Omega}; V)$ (cf.

the "moreover part" of Proposition 3.11). Existence of the μ-measurable selection $\mathscr{H}$ is guaranteed by Lemma 3.41 and optimality of λ follows from Theorem 4.6. □

We conclude the subsection by observing that Theorem 4.6 provides an extra regularity result for the function σ solving the minimization problem (4.5) (the relative stress in the context of elasticity) which *a priori* lies in $L^{p'}_{\mu}(\overline{\Omega}; V)$:

Corollary 4.8. *Assuming that $0 < \mathcal{C}_{\min} < \infty$ let us take any solution λ of the* (FMD) *problem and then fix $\mu = c(\lambda)$ and $\mathscr{H} = \frac{d\lambda}{d\mu} \in L^{\infty}_{\mu}(\overline{\Omega}; \mathcal{K}_1)$.*

Then, each solution σ of the minimization problem (4.5) *for $\mu, \mathscr{H}$ is a function in $L^{\infty}_{\mu}(\overline{\Omega}; V)$ and it is uniform in the sense that*

$$\rho^0(\sigma) = \frac{Z}{C_0} \qquad \mu\text{-a.e.}$$

4.3. Relaxation of the problem ($\mathcal{P}$)

Up to this point we have discussed the (FMD) problem without certainty of existence of its solution λ, which was due to possibility of $\mathcal{C}_{\min} = \infty$. On top of that, even if $\mathcal{C}_{\min}$ is finite, from the pair of problems ($\mathcal{P}$) and ($\mathcal{P}^*$) only the latter is guaranteed to admit a solution. Below we give a list of assumptions that will allow to relax the problem ($\mathcal{P}$) (the smoothness condition on u) and, as a consequence, to assure finiteness of $\mathcal{C}_{\min}$.

By a space of smooth *rigid body motions* (in the nomenclature of elasticity) we shall understand

$$\mathcal{R} := \Big\{ u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U) \ : \ Au = 0 \ \text{ in } \Omega \Big\}.$$

For the linear space $\mathcal{U} := \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U)$ we will choose a norm $|||\cdot|||$ while $\overline{\mathcal{U}}$ will be the Banach space that is a completion of $\mathcal{U}$ with respect to this very norm. The following conditions shall be henceforward enforced:

(B1) $\mathscr{F}(u) = 0$ for any rigid body motion $u \in \mathcal{R}$;
(B2) there exists $C > 0$ such that $|\mathscr{F}(u)| \leq C\,|||u|||$ for any $u \in \mathcal{U}$;
(B3) there exists a linear operator $P_{\mathcal{R}} : \mathcal{U} \to \mathcal{R}$ such that for any $C \geq 0$ the set

$$\Big\{ u - P_{\mathcal{R}} u \ : \ u \in \mathcal{U},\ \|\rho(Au)\|_{\infty,\Omega} \leq C \Big\} \tag{4.8}$$

is $|||\cdot|||$-precompact, i.e. it is precompact in the Banach space $\overline{\mathcal{U}}$.

Remark 4.9. Once the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$) is assumed condition $\|\rho(Au)\|_{\infty,\Omega} \leq C$ in (4.8) may be replaced by $\|Au\|_{\infty,\Omega} \leq C$, cf. Proposition 3.13.

Remark 4.10. It is worth stressing that we do not explicitly make any assumptions on smoothness of the boundary $\partial\Omega$. However, as it will unravel, in order to satisfy the compactness condition (B3) for A being a differential operator regularity of $\partial\Omega$ must be assured so that Morrey's embedding theorem may be employed.

For the set of functions admissible in problem ($\mathcal{P}$)

$$\mathcal{U}_1 := \Big\{ u \in \mathcal{U} = \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U) \ : \ \|\rho(Au)\|_{\infty,\Omega} \leq 1 \Big\} \tag{4.9}$$

by $\overline{\mathcal{U}}_1$ we will denote its closure in the Banach space $\overline{\mathcal{U}}$. Owing to assumption (B2) distribution $\mathscr{F}$ enjoys a unique extension to a linear continuous functional on $\overline{\mathcal{U}}$; we shall preserve the symbol $\mathscr{F} : \overline{\mathcal{U}} \to \mathbb{R}$.

The relaxed variant of problem $(\mathcal{P})$ may be readily posed:

$$\sup\left\{ \mathscr{F}(u) \ : u \in \overline{\mathcal{U}}_1 \right\} \tag{$\overline{\mathcal{P}}$}$$

Theorem 4.11. *Under assumptions* (B1)-(B3) *both problems* $(\overline{\mathcal{P}})$ *and* $(\mathcal{P}^*)$ *have solutions while*

$$Z = \max \overline{\mathcal{P}} = \min \mathcal{P}^* < \infty.$$

As a result there exists solution $\check{\lambda}$ *of* (FMD) *problem for which* $\mathcal{C}_{\min} = \mathcal{C}(\check{\lambda}) < \infty$.

Proof. By continuity of functional $\mathscr{F}$ on $\overline{\mathcal{U}}$ we automatically obtain $\sup(\overline{\mathcal{P}}) = \sup(\mathcal{P}) = \inf(\mathcal{P}^*)$ where the last equality is due to Proposition 4.3. Let us thus take a maximizing sequence $\{u_n\} \subset \mathcal{U}_1$ for the smooth problem $(\mathcal{P})$. By virtue of assumption (B1) the modified sequence $\tilde{u}_n = u_n - P_{\mathcal{R}} u_n \in \mathcal{U}_1$ is another maximizing sequence for $(\mathcal{P})$ since $A\tilde{u}_n = Au_n$. Then, thanks to assumption (B3) the set $\{\tilde{u}_n\}$ is precompact in $\overline{\mathcal{U}}$ thus there exists a subsequence $\tilde{u}_{n_k}$ that converges to $\hat{u} \in \overline{\mathcal{U}}_1$ while $\mathscr{F}(\hat{u}) = \lim_{k\to\infty} \mathscr{F}(\tilde{u}_{n_k}) = \sup(\overline{\mathcal{P}})$ therefore $\hat{u}$ is a solution of $(\overline{\mathcal{P}})$ and $\max(\overline{\mathcal{P}}) < \infty$. Since $Z < \infty$ existence of solution to $(\mathcal{P}^*)$ follows by Proposition 4.3. The second statement follows by combining Proposition 4.1 with Lemma 4.2. □

In the previous subsection Theorem 4.6 established a direct link between solution $\hat{\varsigma}$ of $(\mathcal{P}^*)$ and (in terminology of elasticity) the relative stress σ which for optimal $\check{\lambda}$ solves the stress-based minimization problem (4.5) that dually defines compliance $\mathcal{C}(\check{\lambda})$. Our next effort is to give an analogous result that links solution $\hat{u}$ of $(\overline{\mathcal{P}})$ with solution $\check{u}$ of the maximization problem (4.1). The latter problem requires relaxing the smoothness condition as well – after [37] we propose relaxation that at first glance may appear artificial:

Definition 4.12. Let $\check{\lambda}$ be a solution of the (FMD) problem. The function $\check{u} \in \overline{\mathcal{U}}$ shall be termed a relaxed solution of the maximization problem in (4.1) if $\check{u}$ is a limit in $\overline{\mathcal{U}}$ of a maximizing sequence u_n for problem (4.1) with $\lambda = \check{\lambda}$ and $\{u_n\}$ satisfies the upper bound (4.4).

Proposition 4.13. *Let* $\check{\lambda}$ *be a solution of the* (FMD) *problem where* $0 < \mathcal{C}_{\min} < \infty$. *For a function* $\hat{u} \in \overline{\mathcal{U}}_1$ *we define*

$$\check{u} = \left(\frac{Z}{C_0}\right)^{p'-1} \hat{u}.$$

Then $\hat{u}$ *is a solution of problem* $(\overline{\mathcal{P}})$ *if and only if* $\check{u}$ *is a relaxed solution of problem* (4.1) *for* $\lambda = \check{\lambda}$.

Proof. Let us first assume that $\hat{u}$ solves $(\overline{\mathcal{P}})$. Then there exists a sequence $\hat{u}_n \in \mathcal{U}_1$ such that $|||\hat{u}_n - \hat{u}||| \to 0$. By definition of $\mathcal{U}_1$ we have $\rho(A\hat{u}_n) \leq 1$ and maximizing sequence for problem $\mathcal{C}(\check{\lambda})$ may be chosen as: $\check{u}_n = (Z/C_0)^{p'-1}\,\hat{u}_n$. Indeed, we see that

$$\begin{aligned}\liminf_{n\to\infty}\left\{\mathscr{F}(\check{u}_n) - \int j(\check{\mathscr{H}}, A\check{u}_n)d\check{\mu}\right\} &\geq \liminf_{n\to\infty}\left\{\mathscr{F}(\check{u}_n) - \int \frac{1}{p}\big(\rho(A\check{u}_n)\big)^p d\check{\mu}\right\}\\ \geq \liminf_{n\to\infty}\left\{\mathscr{F}(\check{u}_n) - \int \frac{1}{p}\left(\frac{Z}{C_0}\right)^{p'} d\check{\mu}\right\} &= \lim_{n\to\infty}\left\{\mathscr{F}(\check{u}_n)\right\} - \frac{Z^{\,p'}}{p\,{C_0}^{p'-1}} = \mathcal{C}(\check{\lambda}),\end{aligned}$$

where we have used the fact that $\lim_{n\to\infty}\mathscr{F}(\check{u}_n) = (Z/C_0)^{p'-1}\lim_{n\to\infty}\mathscr{F}(\hat{u}_n) = (Z/C_0)^{p'-1}\,Z$ and that $Z^{p'}/\big(p'{C_0}^{p'-1}\big) = \mathcal{C}_{\min} = \mathcal{C}(\check{\lambda})$ by optimality of $\check{\lambda}$. This shows that $\check{u}_n$ is a maximizing sequence for problem $\mathcal{C}(\check{\lambda})$, which together with $|||\check{u}_n - \check{u}||| = (Z/C_0)^{p'-1}\,|||\hat{u}_n - \hat{u}||| \to 0$ renders $\check{u}$ a relaxed solution of $\mathcal{C}(\check{\lambda})$.

To prove the converse we take a relaxed solution $\check{u}$ of problem $\mathcal{C}(\check{\lambda})$ and we must show that $\hat{u} = \big(\frac{C_0}{Z}\big)^{p'-1}\check{u}$ is a solution for $(\overline{\mathcal{P}})$. It is straightforward that $\hat{u} \in \overline{\mathcal{U}}_1$ based on Definition 4.12 and thus we only have to verify whether $\mathscr{F}(\hat{u}) = Z$. For $\check{u}$ being a relaxed solution of $\mathcal{C}(\check{\lambda})$ there holds the repartition of energy: $\mathscr{F}(\check{u}) = p'\,\mathcal{C}(\check{\lambda})$ (cf. the proof of [37, Proposition 3]). Then, since $\mathcal{C}(\check{\lambda}) = \mathcal{C}_{\min} = Z^{p'}/\big(p'{C_0}^{p'-1}\big)$ we indeed obtain $\mathscr{F}(\hat{u}) = (C_0/Z)^{p'-1}\mathscr{F}(\check{u}) = Z$ and the proof is complete. □

Remark 4.14. Hypothetically, with Theorem 4.6 and Proposition 4.13 at our disposal we recover the relative stress function σ and the displacement function u that solve the elasticity problem in the optimally design body $\lambda = \mathscr{H}\mu$. However, both functions only solve the variational problems (4.5) and (4.1) respectively – we lose the local constitutive relation: $\sigma(x) \in \partial j\big(\mathscr{H}(x), \epsilon(x)\big)$ for μ-a.e. x, where ϵ is the strain function. The issue is that in general u is non-smooth and ϵ cannot be computed as Au. In the forthcoming subsection this gaps gets filled with the aid of μ-tangential calculus: see Corollary 4.24 and the paragraph below it.

4.4. Optimality conditions via the μ-tangential calculus

Below we establish necessary and sufficient conditions for a pair $u \in \overline{\mathcal{U}}$ and $\varsigma \in \mathcal{M}(\overline{\Omega}; V)$ to be optimal for problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$, respectively. The first motivation for deriving the optimality conditions is to efficiently solve the pair of problems analytically. The easiest version of optimality condition is already at our fingertips: based on duality result in Theorem 4.11 a function $u \in \overline{\mathcal{U}}_1$ and a measure $\varsigma \in \mathcal{M}(\overline{\Omega}; V)$ satisfying $A^*\varsigma = \mathscr{F}$ in $\mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U)$ solve problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$, respectively, if and only if

$$\mathscr{F}(u) = \int \rho^0(\varsigma). \tag{4.10}$$

Equation (4.10) or the *global optimality condition* is effective if we are to validate a given competitor (u, ς) as solution. It does not, however, offer a lot of premise on which one can actually guess or find the solution. For this purpose from (4.10) we must extract the local condition linking u and ς or, as it will turn out, between V-valued functions Au and $\sigma = \frac{d\varsigma}{d\mu}\ \ \mu = \rho^0(\varsigma)$-a.e. There is a second reason for deriving the local optimality criteria

that will be fundamental in this thesis: such conditions posed for the optimal membrane problem studied in Chapter 5 will be crucial for constructing the optimal vaulted forms in Chapter 6.

As already mentioned in Remark 4.14 solution $u \in \overline{\mathcal{U}}_1$ of problem $(\overline{\mathcal{P}})$ *a priori* cannot be expected to be smooth therefore to write down the local optimality conditions we must propose a generalized notion of operator A whose domain contains $\overline{\mathcal{U}}_1$. After the original ideas in [35] further developed in [37] to this purpose we shall employ the μ-tangential calculus reviewed in Section 2.5. We must decide on the exponent q: we choose $q = \infty$ which is tailored for the optimal design problems.

Notation and structure of Section 2.5.1 on μ-tangential operator A_μ were designed to match those used in this chapter, up to one exception: below we are bound to assume that $\mu \in \mathcal{M}^b_+(\mathscr{U})$ where $\mathscr{U} := \mathbb{R}^d \setminus \Upsilon_0$. In conjunction with condition on the support of μ being contained in feasible set $\overline{\Omega}$ we may equivalently enforce that: $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\mu(\Upsilon_0) = 0$, which we shall do for the most part of this subsection. As it will appear, condition $\mu(\Upsilon_0) = 0$ is not restrictive in terms of designing the optimal material distribution – it will prove to be a necessary optimality condition, cf. Theorem 4.22.

For any measure $\mu \in \mathcal{M}_+(\overline{\Omega})$ satisfying $\mu(\Upsilon_0) = 0$, for $q = \infty$ and $\mathscr{U} = \mathbb{R}^d \setminus \Upsilon_0$ we may readily build the linear spaces $\overline{\mathcal{G}}^q = \overline{\mathcal{G}}^\infty$ and $\mathcal{N}$ defined at the beginning of Section 2.5.1. To assumptions (B1)-(B3) we add another two conditions:

(B4) for any measure $\mu \in \mathcal{M}_+(\overline{\Omega})$ such that $\mu(\Upsilon_0) = 0$ the stability condition holds:

$$\epsilon \in \mathcal{N}, \quad \psi \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}) \qquad \Rightarrow \qquad \psi\epsilon \in \mathcal{N}, \tag{4.11}$$

where the normal space $\mathcal{N} \subset L^\infty_\mu(\mathbb{R}^d \setminus \Upsilon_0; V)$ is defined in (2.19);

(B5) the energy function j and the cone of admissible anisotropy parameters $\mathcal{K}$ satisfy the ellipticity condition $(\mathrm{A}_{\mathrm{ellip}})$ formulated in Section 3.2.

Thanks to the stability condition (B4) all results from Section 2.5.2 are available, including the following notions: the tangential and normal spaces M_μ and N_μ, respectively, that are μ-measurable multifunctions; the orthogonal projection operator $\Pi_\mu(x) : V \to M_\mu(x)$ defined for μ-a.e. x and finally the μ-tangential operator $A_\mu : \mathrm{dom}(A_\mu) \to L^\infty_\mu(\mathbb{R}^d; M_\mu)$ that for smooth functions $u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; U)$ satisfies $A_\mu u = \Pi_\mu(Au)$.

In the spirit of defining functions j_+, j_- (cf. Section 3.3.2) for μ-a.e. x we introduce the notion of μ-tangential energy function:

$$j_\mu(x, H, \cdot\,) := \big(j^*(H, \cdot\,) + \chi_{M_\mu(x)}(\cdot\,)\big)^*$$

where $\chi_{M_\mu(x)}$ is the indicator function of the linear subspace $M_\mu(x) \subset V$. By proceeding in analogy with the proofs of Propositions 3.20 and 3.26 stated for functions j_+, j_- we arrive at the similar results for j_μ:

Proposition 4.15. *Let $j : \mathcal{K} \times V \to \mathbb{R}$ be any function that satisfies conditions* (A1)-(A4) *(the condition* $(\mathrm{A}_{\mathrm{ellip}})$ *is not obligatory). Then, for μ-a.e. x the function $j_\mu(x, \cdot, \cdot) : \mathcal{K} \times V \to \mathbb{R}_+$ satisfies* (A1)-(A4) *as well, whilst*

$$j_\mu(x, H, \xi) = \inf_{\zeta \in N_\mu(x)} j(H, \xi + \zeta), \qquad j^*_\mu(x, H, \sigma) = j^*(H, \sigma) + \chi_{M_\mu(x)}(\sigma) \tag{4.12}$$

where $N_\mu(x) = \big(M_\mu(x)\big)^\perp$. *Consequently, for* μ*-a.e.* x, *we may define functions* $\bar{j}_\mu(x,\,\cdot\,) : V \to \mathbb{R}_+$, $\rho_\mu(x,\,\cdot\,) : V \to \mathbb{R}_+$ *and sets* $\bar{\mathcal{K}}_{1,\mu}(x,\,\cdot\,)$, $\underline{\mathcal{K}}_{1,\mu}(x,\,\cdot\,)$ *in accordance with Section 3.2. The following formulas hold true:*

$$\rho_\mu(x,\xi) = \inf_{\zeta \in N_\mu(x)} \rho(\xi+\zeta), \qquad \rho^0_\mu(x,\sigma) = \rho^0(\sigma) + \chi_{M_\mu(x)}(\sigma) \tag{4.13}$$

and

$$\underline{\mathcal{K}}_{1,\mu}(x,\sigma) = \begin{cases} \underline{\mathcal{K}}_1(\sigma) & \text{if} \quad \sigma \in M_\mu(x), \\ \mathcal{K}_1 & \text{if} \quad \sigma \notin M_\mu(x). \end{cases}$$

Let us note that the first formula in (4.12) was originally given as the definition of j_μ in the original work [40]. At μ-a.e. x the result above unlocks all the tools developed in Section 3.2. The statement (ii) of Theorem 3.14 will be of particular importance below. Thanks to formulas from Proposition 4.15 the latter result can be slightly rewritten by disposing of the lower index μ at strategic places:

Corollary 4.16. *Let* $j : \mathcal{K} \times V \to \mathbb{R}$ *satisfy conditions* (A1)-(A4) *(the condition* $(\mathrm{A}_{\mathrm{ellip}})$ *is not obligatory). For* μ*-a.e.* x *and for any triple* $(\xi, \sigma, \breve{H}) \in V \times V \times \mathcal{K}$ *satisfying:*

$$\rho_\mu(x,\xi) \leq 1, \qquad \sigma \neq 0, \quad \rho^0(\sigma) < \infty, \quad \sigma \in M_\mu(x), \qquad \breve{H} \in \mathcal{K}_1$$

the following conditions are equivalent

(1) $\langle \xi, \sigma \rangle = \rho^0(\sigma)$ *and* $\breve{H} \in \underline{\mathcal{K}}_1(\sigma)$,
(2) $\frac{1}{\rho^0(\sigma)}\,\sigma \in \partial j_\mu\big(x, \breve{H}, \xi\big)$.

Moreover, for each of the conditions (1), (2) to be true it is necessary that $\rho_\mu(x,\xi) = 1$;

Proof. The point of departure is the assertion (ii) in Theorem 3.14 written down for $j = j_\mu(x,\,\cdot\,,\,\cdot\,)$. We replace the original condition $\rho^0_\mu(x,\sigma) < \infty$ by the pair: $\rho^0(\sigma) < \infty$, $\sigma \in M_\mu(x)$. Consequently, owing to Proposition 4.15, we may write $\rho^0(\sigma)$ and $\underline{\mathcal{K}}_1(\sigma)$ in place of $\rho^0_\mu(x,\sigma)$ and $\underline{\mathcal{K}}_{1,\mu}(x,\sigma)$, respectively. □

With the use of A_μ and j_μ a relaxation of the maximization problem in (4.1) may be proposed as an alternative to relaxation in accordance with Definition 4.12:

Proposition 4.17. *Let* $\lambda = \mathscr{H}\mu \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ *where* $\mu \in \mathcal{M}_+(\overline{\Omega})$, $\mathscr{H} \in L^\infty_\mu(\overline{\Omega}; \mathcal{K})$ *and moreover* $\mu(\Upsilon_0) = 0$ *holds. Then* $(x,\xi) \mapsto j_\mu\big(x, \mathscr{H}(x), \xi\big)$ *is a normal convex integrand and a Carathéodory function while*

$$\mathcal{C}(\lambda) = \sup\left\{ \mathscr{F}(u) - \int j_\mu\big(x, \mathscr{H}, A_\mu u\big)\, d\mu \;:\; u \in \mathrm{dom}(A_\mu) \right\}. \tag{4.14}$$

Moreover, $u \in \mathrm{dom}(A_\mu)$ *is a solution of the problem above if and only if there exists* $\sigma \in L^{p'}_\mu(\overline{\Omega}; V)$ *such that*

$$A^*(\sigma\mu) = \mathscr{F} \quad \text{in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U), \qquad \sigma \in \partial j_\mu\big(x, \mathscr{H}, A_\mu u\big) \quad \mu\text{-a.e.} \tag{4.15}$$

Proof. According to Proposition 3.36 the function $(x,\xi) \mapsto j\big(\mathscr{H}(x), \xi\big)$ is a Carathéodory integrand hence it is a normal convex integrand by virtue of [137, Lemma 2] and thus such is the convex conjugate $(x,\sigma) \mapsto j^*\big(\mathscr{H}(x), \sigma\big)$, cf. [137, Lemma 5]. Since M_μ is a μ-measurable

multifunction it is straightforward to check that $(x,\sigma)\mapsto\chi_{M_\mu(x)}(\sigma)$ is a normal convex integral as well. By [139, Corollary 4.2] the sum $(x,\sigma)\mapsto j^*\big(\mathscr{H}(x),\sigma\big)+\chi_{M_\mu(x)}(\sigma)$ is also a normal convex integrand and finally such is $(x,\xi)\mapsto j_\mu\big(x,\mathscr{H}(x),\xi\big)$ as its convex conjugate.

By representation (4.12) for μ-a.e. x there holds $j_\mu\big(x,\mathscr{H}(x),\,\cdot\,\big)\le j\big(\mathscr{H}(x),\,\cdot\,\big)$ therefore the convex functional $\epsilon\mapsto\int j_\mu\big(x,\mathscr{H},\epsilon\big)\,d\mu$ is continuous on $L^p_\mu(\overline{\Omega};V)$ based on the same argument as in Corollary 3.37 while its convex conjugate equals $L^{p'}_\mu(\overline{\Omega};V)\ni\sigma\mapsto\int j^*_\mu\big(x,\mathscr{H},\sigma\big)\,d\mu$. By repeating the arguments from the proof of Proposition 4.4 we obtain

$$\widetilde{\mathcal{C}}:=\sup_{u\in\mathcal{U}}\left\{\mathscr{F}(u)-\int j_\mu\big(x,\mathscr{H},Au\big)\,d\mu\right\}=\inf_{\sigma\in L^{p'}_\mu(\overline{\Omega};V)}\left\{\int j^*_\mu\big(\mathscr{H},\sigma\big)\,d\mu\ :\ A^*(\sigma\mu)=\mathscr{F}\right\}$$

Then, by claim (i) of Theorem 2.57 the distributional equation $A^*(\sigma\mu)=\mathscr{F}$ implies that $\sigma\in M_\mu$ μ-a.e. therefore $j^*_\mu\big(\mathscr{H},\sigma\big)=j_\mu\big(\mathscr{H},\sigma\big)$ and $\widetilde{\mathcal{C}}=\mathcal{C}(\lambda)$ by Proposition 4.4.

For each $u\in\mathrm{dom}(A_\mu)$ there exists by definition a sequence $\{u_n\}\subset\mathcal{U}$ and $\epsilon_0\in L^p_\mu(\overline{\Omega};V)$ with $\epsilon_0(x)\in N_\mu(x)$ μ-a.e. such that $|||u_n-u|||\to0$ and $Au_n\to A_\mu u+\epsilon_0$ in $L^p_\mu(\overline{\Omega};V)$ (see Remark 2.47). From (4.12) we see that $j_\mu\big(x,\mathscr{H},A_\mu u+\epsilon_0\big)=j_\mu\big(x,\mathscr{H},A_\mu u\big)$ μ-a.e. and the asserted equality (4.14) follows by continuity of $\mathscr{F}$ and $\int j_\mu\big(x,\mathscr{H},\,\cdot\,\big)\,d\mu$.

We readily see that functions $u\in\mathrm{dom}(A_\mu)$ and $\sigma\in L^{p'}_\mu(\overline{\Omega};V)$ solve (4.14) and the minimization problem above if and only if $A^*(\sigma\mu)=\mathscr{F}$ and $\mathscr{F}(u)-\int j_\mu\big(x,\mathscr{H},A_\mu u\big)\,d\mu=\int j^*_\mu\big(\mathscr{H},\sigma\big)\,d\mu$. By exploiting the integration by parts formula in Theorem 2.57 the latter condition may be rewritten as constitutive law $\langle A_\mu,\sigma\rangle=j_\mu\big(x,\mathscr{H},A_\mu u\big)+j^*_\mu\big(\mathscr{H},\sigma\big)$ μ-a.e. and the "moreover part" of the claim follows. □

Remark 4.18. In the pioneering work [40], for $A=\nabla$ and $q=p$ rather than $q=\infty$, the operator $A_\mu=\nabla_\mu$ and its domain $\mathrm{dom}(A_\mu)=W^{1,p}_\mu$ were designed to compute relaxation of the functional $\mathcal{D}(\mathbb{R}^d)\ni u\mapsto\int j(\mathscr{H},\nabla u)\,d\mu$, i.e. the lower semi-continuous envelope of this functional extended to $W^{1,p}_\mu$ by ∞. Under additional assumption on ellipticity (which here can be translated to "ellipticity" of function $\mathscr{H}$) the relaxed functional was established to be precisely $W^{1,p}_\mu\ni u\mapsto\int j_\mu(x,\mathscr{H},\nabla_\mu u)\,d\mu$. This allowed the authors to show existence of problem (4.14) provided that $\mathscr{F}$ is a functional in $\big(L^p_\mu(\mathbb{R}^d)\big)^*$. Herein we cannot assume the suitable ellipticity of $\mathscr{H}$ while functional $\mathscr{F}$ may e.g. be a Radon measure. Nevertheless, we shall see that solution of (4.14) exists once μ and $\mathscr{H}$ are chosen optimally, cf. Corollary 4.26 below.

The following result is a slight modification of [35, Lemma 3.6]; it ensures that any solution u of the relaxed problem $(\overline{\mathcal{P}})$ is an element of the space $\mathrm{dom}(A_\mu)$ and that $A_\mu u$ is meaningful as a result:

Lemma 4.19. *For $\mu\in\mathcal{M}_+(\overline{\Omega})$ satisfying $\mu(\Upsilon_0)=0$ under assumptions* (B1)-(B5) *there holds the inclusion*

$$\overline{\mathcal{U}}_1\subset\mathrm{dom}(A_\mu).$$

Moreover, for any $u\in\overline{\mathcal{U}}_1$

$$\rho_\mu\big(x,A_\mu u\big)\le1\qquad\text{for }\mu\text{-a.e. }x.$$

Proof. Let u be an element of $\mathcal{U}_1$. Then, by definition there exists a sequence $\{u_n\} \subset \mathcal{U}$ such that $\|\rho(Au_n)\|_{\infty,\Omega} \leq 1$ and $\|\!|u_n - u|\!\| \to 0$. Thanks to ($\mathrm{A}_{\mathrm{ellip}}$) we also obtain $\|Au_n\|_{\infty,\Omega} \leq 1$ for each n (cf. Proposition 3.13) and thus also $\|A_\mu u_n\|_{L^\infty_\mu} \leq 1$. The fact that $u \in \mathrm{dom}(A_\mu)$ follows directly by Corollary 2.55. In particular, due to the suitable pre-compactness of the sequence $\{Au_n\}$, up to extracting a subsequence we have $Au_n \overset{*}{\rightharpoonup} \epsilon$ in $L^\infty_\mu(\overline{\Omega}; V)$ for some ϵ. By definition of the operator A_μ there must hold $A_\mu u = \Pi_\mu \epsilon$ or equivalently $\epsilon = A_\mu \mu + \zeta$ where $\zeta \in N_\mu$ μ-a.e. For any Borel set $B \subset \overline{\Omega}$ we may write down the chain of inequalities:

$$\int_B \rho_\mu(x, A_\mu u)\, d\mu \leq \int_B \rho(A_\mu u + \zeta)\, d\mu \leq \liminf_n \int_B \rho(Au_n)\, d\mu \leq \int_B d\mu$$

where the first inequality is due the first formula in (4.12) and to obtain the second one we acknowledge that for any $q \in [1, \infty)$

- Au_n converges to $\epsilon = A_\mu u + \zeta$ weakly in $L^q_\mu(\overline{\Omega}; V)$;
- the functional $L^q_\mu(\overline{\Omega}; V) \ni \epsilon \mapsto \int_B \rho(\epsilon)\, d\mu$ is continuous for the norm topology owing to Proposition 2.6 – indeed, ρ is a convex function that can be estimated from above by $C|\cdot|$, see Proposition 3.13;
- consequently the same functional is lower semi-continuous for the weak topology on $L^q_\mu(\overline{\Omega}; V)$, see e.g. [71, Chapter 3, Theorem 1.2].

The moreover part readily follows due to arbitrariness of the Borel set B. □

We can readily give the necessary and sufficient optimality conditions for the pair of problem $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$; the result below merges [35, Theorem 3.9] and [37, Theorem 3]:

Theorem 4.20 (**Bouchitté-Buttazzo-Fragalà optimality conditions**). *Let us take a triple (u, μ, σ) such that: $u \in \overline{\mathcal{U}}$, $\mu \in \mathcal{M}_+(\overline{\Omega})$, $\sigma \in L^\infty_\mu(\overline{\Omega}; V)$ with $\rho^0(\sigma) = 1$ μ-a.e. The function u and measure $\varsigma = \sigma\mu$ solve problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$, respectively, if and only if the following optimality conditions are met:*

$$\begin{cases} (i) & A^*(\sigma\mu) = \mathscr{F} \quad \text{in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U), \\ (ii) & u \in \overline{\mathcal{U}}_1, \\ (iii) & \mu(\Upsilon_0) = 0, \\ (iv) & \langle A_\mu u, \sigma \rangle = 1 \quad \mu\text{-a.e.} \end{cases} \tag{4.16}$$

Proof. Conditions (i) and (ii) are the admissibility criteria for $\varsigma = \sigma\mu$ and u in problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$, respectively, hence they will be assumed throughout the proof. By virtue of the zero-gap result in Proposition 4.3 the inequality

$$\mathscr{F}(u) \leq \int_{\mathbb{R}^d} \rho^0(\varsigma) = \int_{\mathbb{R}^d} \rho^0(\sigma)\, d\mu$$

becomes an equality if and only if u and ς solve $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$ respectively.

Let us put $\mu_0 = \mu \llcorner (\mathbb{R}^d \setminus \Upsilon_0)$. Since $u \in \mathrm{dom}(A_{\mu_0})$ by Lemma 4.19, based on (i) we may exploit Theorem 2.57 to infer that $\sigma \in M_{\mu_0}$ μ_0-a.e. and to apply the integration by parts formula in Theorem 2.57 thus rewriting the inequality above as

$$\int_{\mathbb{R}^d \setminus \Upsilon_0} \langle A_{\mu_0} u, \sigma \rangle\, d\mu_0 \leq \int_{\mathbb{R}^d} \rho^0(\sigma)\, d\mu.$$

Moreover, from Lemma 4.19 we learn that $\rho_{\mu_0}(x, A_{\mu_0}u) \leq 1$ μ_0-a.e. and therefore $\langle A_{\mu_0}u, \sigma\rangle \leq \rho^0_{\mu_0}(x,\sigma) = \rho^0(\sigma)$, where the latter equality follows by the second formula in (4.13). Hence, upon recalling that $\rho^0(\sigma) = 1$ μ-a.e., u and $\varsigma = \sigma\mu$ are solutions if and only if conditions (iii) and $\langle A_{\mu_0}u, \sigma\rangle = 1$ hold μ_0-a.e., but in the latter condition μ_0 may be replaced by μ due to (iii). □

Remark 4.21. From the proof one can observe that condition $\rho^0(\sigma) = 1$ μ-a.e. may be dropped if we replace 1 on the RHS of condition (iv) in (4.16) by $\rho^0(\sigma)$. Then we may consider any pair $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\sigma \in L^1_\mu(\overline{\Omega}; V)$.

In the case of works [35] and [37], both on mass optimization problem, the route from optimality conditions for the pair $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$ to conditions for optimality of μ in (MOP) is straightforward through the result analogous to Theorem 4.6 herein. In the case of the (FMD) problem we must additionally handle optimality of the anisotropy function $\mathscr{H}$. By virtue of Theorem 4.6 this issue gets fixed if we append the optimality criteria above with the fifth condition: $\mathscr{H}(x) \in \underline{\mathcal{K}}_1\big(\sigma(x)\big)$ μ-a.e., which in principal requires solving the problem $\min_{H\in\mathcal{K}_1} j^*\big(H, \sigma(x)\big)$ for μ-a.e. x. However, wishing to keep the spirit of optimality criteria that do not entail solving any min or max problems, we exploit Corollary 4.16 to propose an adaptation of Bouchitté-Buttazzo-Fragalà result towards the (FMD) problem:

Theorem 4.22. *Let us take a quadruple $(u, \mu, \sigma, \mathscr{H})$ such that: $u \in \overline{\mathcal{U}}$, $\mu \in \mathcal{M}_+(\overline{\Omega})$, $\sigma \in L^\infty_\mu(\overline{\Omega}; V)$, $\mathscr{H} \in L^\infty_\mu(\overline{\Omega}; \mathcal{K})$ and $\rho^0(\sigma) = 1$, $\mathscr{H} \in \mathcal{K}_1$ μ-a.e.*

The function u and measure $\varsigma = \sigma\mu$ solve problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$, respectively, and moreover $\mathscr{H}(x) \in \underline{\mathcal{K}}_1\big(\sigma(x)\big)$ μ-a.e. if and only if the following optimality conditions are met:

$$\begin{cases} (i) & A^*(\sigma\mu) = \mathscr{F} \quad \text{in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U), \\ (ii) & u \in \overline{\mathcal{U}}_1, \\ (iii) & \mu(\Upsilon_0) = 0, \\ (iv) & \sigma \in \partial j_\mu\big(x, \mathscr{H}, A_\mu u\big) \quad \mu\text{-a.e.} \end{cases} \tag{4.17}$$

Proof. Clearly the quadruple is optimal if and only if conditions (i)-(iv) from Theorem 4.20 are met and moreover $\mathscr{H}(x) \in \underline{\mathcal{K}}_1\big(\sigma(x)\big)$ μ-a.e. By virtue of the claim (i) in Theorem 2.57 the condition (i) above in (4.16) implies that $\sigma \in M_\mu$ μ-a.e., while by Lemma 4.19 we know that (ii) in (4.16) guarantees that $\rho_\mu(x, A_\mu u) \leq 1$ μ-a.e. Therefore, for μ-a.e. x, the prerequisites of Corollary 4.16 are met for the triple $\xi = A_\mu u(x)$, $\sigma(x)$, $\breve{H} = \mathscr{H}(x)$ and readily the pair of conditions $\langle (A_\mu u)(x), \sigma(x)\rangle = 1$ and $\mathscr{H}(x) \in \underline{\mathcal{K}}_1\big(\sigma(x)\big)$ may be equivalently put as (iv) in the assertion above (note that $\rho^0(\sigma) = 1$). □

Remark 4.23. If u is smooth, i.e. $u \in \mathcal{U}$, engaging the μ-tangential tools may be dropped and condition (iv) above can be replaced by

(iv') $\sigma \in \partial j\big(\mathscr{H}, Au\big)$ μ-a.e.

Indeed, first we observe that $\langle A_\mu u, \sigma\rangle = \langle \Pi_\mu(Au), \sigma\rangle = \langle Au, \sigma\rangle$ and $j_\mu(x, \mathscr{H}, A_\mu u) \leq j(\mathscr{H}, Au) \leq 1/p$ μ-a.e. furnishing implication (iv') $\Rightarrow$ (iv) by virtue of (3.5). The converse

follows from the fact that (iv) implies that $j_\mu(x, \mathscr{H}, A_\mu u) = 1/p$ μ-a.e., cf. the proof of Theorem 3.14 for $j = j_\mu$.

For a quadruple $(u, \mu, \sigma, \mathscr{H})$ satisfying prerequisites of Theorem 4.22 under assumption that $0 < Z < \infty$ we define: $\check{u} = (Z/C_0)^{p'-1}u$, $\check{\sigma} = Z/C_0\,\sigma$, $\check{\mu} = C_0/Z\,\mu$, $\check{\mathscr{H}} = \mathscr{H}$ to discover that

$$\langle A_{\check{\mu}}\check{u}, \check{\sigma}\rangle = \left(\frac{Z}{C_0}\right)^{p'-1}\frac{Z}{C_0}\,\langle A_\mu u, \sigma\rangle, \qquad j_{\check{\mu}}\big(x, \check{\mathscr{H}}, A_{\check{\mu}}\check{u}\big) = \left(\left(\frac{Z}{C_0}\right)^{p'-1}\right)^{p} j_\mu\big(x, \mathscr{H}, A_\mu u\big)$$

$$j^*_{\check{\mu}}\big(x, \check{\mathscr{H}}, \check{\sigma}\big) = \left(\frac{Z}{C_0}\right)^{p'} j^*_\mu\big(x, \mathscr{H}, \sigma\big)$$

By observing that $(p'-1)\,p = p'$ and acknowledging (3.5) we obtain equivalence of conditions $\sigma \in \partial j_\mu\big(x, \mathscr{H}, A_\mu u\big)$ and $\check{\sigma} \in \partial j_{\check{\mu}}\big(x, \check{\mathscr{H}}, A_{\check{\mu}}\check{u}\big)$. Through Theorem 4.6 the optimality criteria in Theorem 4.22 may be translated to the original framework of the (FMD) problem:

Corollary 4.24. *Assuming that $0 < Z < \infty$ let us take a quadruple $(u, \mu, \sigma, \mathscr{H})$ such that: $u \in \overline{\mathcal{U}}$, $\mu \in \mathcal{M}_+(\overline{\Omega})$, $\sigma \in L^{p'}_\mu(\overline{\Omega}; V)$, $\mathscr{H} \in L^\infty_\mu(\overline{\Omega}; \mathcal{K})$ and $\mathscr{H} \in \mathcal{K}_1$ μ-a.e.*

The consitutive field $\lambda = \mathscr{H}\mu$ solves the (FMD) *problem, the function u is a relaxed solution of $\mathcal{C}(\lambda)$ and the function σ solves problem* (4.5) *if and only if*

$$\begin{cases} (i) & A^*(\sigma\mu) = \mathscr{F} \quad \text{in } \mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; U), \\ (ii) & u \in \left(\frac{Z}{C_0}\right)^{p'-1}\overline{\mathcal{U}}_1, \\ (iii) & \mu(\Upsilon_0) = 0, \\ (iv) & \sigma \in \partial j_\mu\big(x, \mathscr{H}, A_\mu u\big) \quad \mu\text{-a.e.}, \\ (v) & \rho^0(\sigma) = \frac{Z}{C_0} \quad \mu\text{-a.e.} \end{cases} \tag{4.18}$$

Remark 4.25. It is worth noting that the admissibility condition $\int c(\lambda) \le C_0$ is not present in Corollary 4.24 above. Instead it implicitly follows from conditions (i)-(v) above.

Speaking in terms of elasticity, conditions (i) and (iv) in Corollary 4.24 constitute a system that renders the pair (u, σ) a solution of elasticity problem of a body with elastic material distribution μ and anisotropy field $\mathscr{H}$: the relative stress function σ solves the equilibrium equation (i), whilst the μ-tangential strain function $A_\mu u$ complies with the constitutive law (iv) for this very relative stress σ. Ultimately, due to Proposition 4.17, we observe that (i), (iv) ensure optimality of u in the displacement-based variational elasticity problem (4.14), more precisely:

Corollary 4.26. *Let us assume that $\lambda = \mathscr{H}\mu$ is a solution of* (FMD) *problem with $c(\mathscr{H}) = 1$ μ-a.e. For a function $u \in \overline{\mathcal{U}}$ let us assume that one of the two equivalent conditions holds:*

(i) u is a relaxed solution of maximization problem (4.1) *for λ;*
(ii) $(C_0/Z)^{p'-1}\,u$ solves problem $(\overline{\mathcal{P}})$.

Then u solves the μ-tangential maximization problem (4.14) *for $\mu, \mathscr{H}$.*

4.5. The three main frameworks of the Free Material Design problem

In the present section we give three specifications of the general theory of Free Material Design put forward in this chapter. First we shall investigate the Free Material Design of 2D or 3D elastic bodies modelled by the classical elasticity theory, i.e. where A is the symmetric part of the gradient. Precisely in this context the (FMD) problem was originally posed in [136] and developed in [20], [21] (for the AMD setting) and then more recently in e.g. [64], [62], [69] (for AMD and IMD settings and others). For the second framework we will choose the Kirchhoff plate theory, that is we fix A as the second order Hessian operator ∇^2. The Free Material Design of Kirchhoff plates in the AMD setting was tackled numerically in [169] whilst their (MOP) problem was the main topic in [37]. Finally we will localize the theory of (FMD) for the optimal design of heat conductor where the state function u is scalar.

4.5.1. Framework #1: Free Material Design of 2D and 3D elastic bodies

For $d = 2$ or $d = 3$ we choose a feasible bounded domain $\Omega \subset \mathbb{R}^d$; we assume that the boundary $\partial\Omega$ is Lipschitz continuous. We consider the Free Material Design problem (FMD) in the following framework:

$$\begin{cases} U = \mathbb{R}^d, \quad \mathcal{U} = \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d), \\ |||\cdot||| = \|\cdot\|_{\infty,\Omega}, \\ \mathscr{F}(u) = \int \langle u, F\rangle \quad \text{where } F \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d), \\ V = \mathcal{S}^{d\times d}, \quad Au = e(u) := \frac{1}{2}\Big(\nabla u + (\nabla u)^\top\Big), \\ j : \mathcal{K} \times \mathcal{S}^{d\times d} \to \mathbb{R}_+ \quad \text{satisfying (A1)-(A4) and } (\mathrm{A}_{\mathrm{ellip}}). \end{cases} \tag{4.19}$$

Readily, for a given constitutive field $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ we recover the elastic compliance

$$\mathcal{C}(\lambda) = \sup\left\{ \int \langle u, F\rangle - \int j\big(\lambda, e(u)\big) \ : \ u \in \mathcal{D}\big(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d\big) \right\} \tag{4.20}$$

of an elastic body λ that is kinematically fixed on Υ_0 and is subject to the load F being a vector valued force field. Recalling that

$$\mathcal{R} = \Big\{ u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d) \ : \ e(u) = 0 \ \text{ in } \Omega \Big\},$$

henceforward we shall assume that one of the two scenarios occur:

(1) Υ_0 is an empty set (no Dirichlet conditions) and the load F is *balanced*, i.e.

$$\int_{\overline{\Omega}} F = 0, \qquad \int_{\overline{\Omega}} x \wedge F = 0 \tag{4.21}$$

where $\wedge$ stands for the cross vector product. In this case we choose the linear operator $P_{\mathcal{R}} : \mathcal{U} \to \mathcal{R}$ that satisfies

$$(P_{\mathcal{R}} u)(x) = a_0(u) + A_0(u)\, x$$

where $A_0(u) = \fint_\Omega \frac{1}{2}\big(\nabla u - (\nabla u)^\top\big)\, d\mathcal{L}^d$ and $a_0(u) = \fint_\Omega \big(u(x) - A_0(u)\, x\big)\, d\mathcal{L}^d$

(2) or the affine hull of Υ_0 contains a $(d-1)$-dimensional affine space while $F \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ is arbitrary. In this case $\mathcal{R} = \{0\}$ and thus we set

$$P_{\mathcal{R}} u = 0.$$

Remark 4.27. In the case (1) the function $P_{\mathcal{R}} u$ in general does not have a compact support while in principal $\mathcal{R} \subset \mathcal{D}(\mathbb{R}^d; \mathbb{R}^d)$. In fact, for further purposes we may take an operator $\widetilde{P}_{\mathcal{R}} u = \varphi\, P_{\mathcal{R}} u$ instead, where $\varphi \in \mathcal{D}(\mathbb{R}^d)$ is any cut-off function satisfying $\varphi \equiv 1$ on Ω.

Let us define a linear space of functions over Ω:

$$\mathcal{R}_\Omega = \Big\{ u = u(x) = a_0 + A_0 x \ : \ a_0 \in \mathbb{R}^d,\ A_0 \in \mathbb{R}^{d\times d}_{\mathrm{skew}},\ \ u = 0 \text{ on } \Upsilon_0 \Big\},$$

namely any function $u \in \mathcal{R}_\Omega$ consists of translation a_0 and an infinitesimal rotation represented by a skew-symmetric matrix A_0; in addition a homogeneous Dirichlet condition is imposed on Υ_0. Then it can be easily checked that in each case (1) or (2) the space $\mathcal{R}_\Omega$ is precisely equal to $\big\{ u|_\Omega : u \in \mathcal{R} \big\}$. Moreover, for any $q > d$ we observe that

$$\mathcal{R}_\Omega = \Big\{ W^{1,q}(\Omega; \mathbb{R}^d) \ : \ e(u) = 0,\ \ u = 0 \text{ on } \Upsilon_0 \Big\} \tag{4.22}$$

(the lower bound for q is to guarantee that condition $u = 0$ on Υ_0 is well defined) and $P_{\mathcal{R}} : W^{1,q}(\Omega; \mathbb{R}^d) \to \mathcal{R}_\Omega$ is a projection operator, i.e. $P_{\mathcal{R}} u = u$ for every $u \in \mathcal{R}_\Omega$. On top of that, $P_{\mathcal{R}}$ is continuous in the weak topology of $W^{1,q}(\Omega; \mathbb{R}^d)$.

Remark 4.28. One can observe that conditions (1), (2) are restrictive in the sense that they both rule out e.g. a one-point Dirichlet region $\Upsilon_0 = \{x_0\}$. For such set definition $\mathcal{R} = \big\{ u \in \mathcal{D}\big(\mathbb{R}^d \setminus \{x_0\}; \mathbb{R}^d\big) \ : e(u) = 0 \ \text{ in } \Omega \big\}$ is not adhering to the classical understanding of the space of rigid body motions $\mathcal{R}_\Omega$: we have $\mathcal{R} = \{0\}$ while any rotation $u(x) = A_0(x - x_0)$ is in fact a rigid body motion. In order to handle the general case of Υ_0 one needs to *a priori* define $\mathcal{R} = \big\{ u \in \mathcal{D}\big(\mathbb{R}^d; \mathbb{R}^d\big) \ : \ e(u) = 0 \ \text{ in } \Omega,\ u = 0 \text{ on } \Upsilon_0 \big\}$ as was proposed in [35]. This way restriction of $\mathcal{R}$ to Ω is always equal to $\mathcal{R}_\Omega$ (which is essential for the proof of Proposition 4.29 below). Such definition of $\mathcal{R}$, however, does not fall into the general framework of (FMD) problem herein given. Its adaptation to the new $\mathcal{R}$ is possible but entails some subtle technical difficulties, see e.g. [35, Lemma 3.6]. For those reasons, the cases covered by scenarios (1) and (2) were deemed by the author sufficient for the purposes of this thesis.

We shall now check assumptions (B1)-(B5). In any scenario (1) or (2) it is straightforward to verify assumption (B1), i.e. $\int \langle u, F \rangle = 0$ for any $u \in \mathcal{R}$. Validity of (B2) is clear since F is a Radon measure and $|||\cdot|||$ is the uniform norm. Next, the μ-tangential calculus for the operator $A = e$ has been already handled in Section 2.5.3, including checking the stability assumption (B4) (referred to as $(\mathrm{H}_{\mathrm{stab}})$ therein). Assumption (B5) holds trivially, i.e. $(\mathrm{A}_{\mathrm{ellip}})$ is explicitly enforced in (4.19). It thus remains to verify the compactness assumption (B3). Since $|||\cdot||| = \|\cdot\|_{\infty,\Omega}$ we can see that condition (B3) is equivalent to the result below:

Proposition 4.29. *Assuming that the boundary $\partial\Omega$ is Lipschitz continuous and* ($\mathrm{A}_{\mathrm{ellip}}$) *holds, for any $C > 0$ the set*

$$\Big\{u - P_{\mathcal{R}}u \,:\, u \in \mathcal{D}(\mathbb{R}^d;\mathbb{R}^d),\ \|\rho\big(e(u)\big)\|_{\infty,\Omega} \le C\Big\} \tag{4.23}$$

is pre-compact in $C(\overline{\Omega};\mathbb{R}^d)$.

Proof. For any $q \in (d,\infty)$ we will show that there exists a constant $C_1 > 0$ such that a Poincaré-type inequality holds:

$$\|u - P_{\mathcal{R}}u\|_{L^q(\Omega)} \le C_1\,\|e(u)\|_{L^q(\Omega)} \qquad \forall\, u \in \mathcal{D}(\mathbb{R}^d;\mathbb{R}^d). \tag{4.24}$$

We shall argue by contradiction assuming that this is not the case. Then there exists a sequence $\{u_n\}$ such that $\|\tilde{u}_n\|_{L^q(\Omega)} = 1$ and $\|e(\tilde{u}_n)\|_{L^q(\Omega)} \to 0$, where $\tilde{u}_n = u_n - P_{\mathcal{R}}u_n$ and thus $e(\tilde{u}_n) = e(u_n)$. From Korn's inequality we have $\|\nabla\tilde{u}_n\|_{L^q(\Omega)} \le C_2\big(\|\tilde{u}_n\|_{L^q(\Omega)} + \|e(\tilde{u}_n)\|_{L^q(\Omega)}\big)$ for $C_2 > 0$. The sequence $\{\tilde{u}_n\}$ is therefore bounded in $W^{1,q}(\Omega)$ and is pre-compact in $C(\overline{\Omega};\mathbb{R}^d)$ by Morrey's embedding theorem. Since $W^{1,q}(\Omega;\mathbb{R}^d)$ is reflexive there exists a function $\tilde{u} \in W^{1,q}(\Omega;\mathbb{R}^d)$ such that (up to choosing subsequence) $\tilde{u}_n \to \tilde{u}$ uniformly and $\nabla\tilde{u}_n \rightharpoonup \nabla\tilde{u}$ weakly in $L^q(\Omega;\mathbb{R}^{d\times d})$. This implies that $\|\tilde{u}\|_{L^q(\Omega)} = 1$, moreover $u = 0$ on Υ_0 and $e(\tilde{u}_n) \rightharpoonup e(\tilde{u}) = 0$ since $\|e(\tilde{u}_n)\|_{L^q(\Omega)} \to 0$ and we infer that $\tilde{u} \in \mathcal{R}_\Omega$ thanks to equality (4.22). Then, since $P_{\mathcal{R}}$ as an operator on $W^{1,q}(\Omega;\mathbb{R}^d)$ is a linear projection onto $\mathcal{R}_\Omega$, we notice that $P_{\mathcal{R}}\tilde{u}_n = P_{\mathcal{R}}u_n - P_{\mathcal{R}}(P_{\mathcal{R}}u_n) = 0$ in Ω and, since $P_{\mathcal{R}}$ is continuous in weak topology of $W^{1,q}(\Omega;\mathbb{R}^d)$, we infer that $P_{\mathcal{R}}\tilde{u} = 0$ and thus necessarily $\tilde{u} = 0$, which contradicts $\|\tilde{u}\|_{L^q(\Omega)} = 1$.

Let us now consider any function $\tilde{u} = u - P_{\mathcal{R}}u$ in the set (4.23). Combining (4.24) with Korn's inequality for any $q \in (d,\infty)$ we obtain $\|\tilde{u}\|_{W^{1,q}(\Omega)} \le C_3\|e(u)\|_{L^q(\Omega)}$ while $\|e(u)\|_{L^q(\Omega)} \le C_4\|\rho\big(e(u)\big)\|_{L^q(\Omega)}$ owing to assumption ($\mathrm{A}_{\mathrm{ellip}}$) (cf. Proposition 3.13). The set (4.23) is therefore bounded in $W^{1,q}(\Omega;\mathbb{R}^d)$ for any $q \in (d,\infty)$ and the assertion follows by Morrey's embedding theorem. □

With conditions (B1)-(B5) met we have all the result from Sections 4.1 - 4.4 at our disposal. Thanks to Theorem 4.11 we know that the (FMD) problem attains solution λ. Below we present the pair of problems ($\mathcal{P}$) and ($\mathcal{P}^*$) corresponding to optimal design of elastic bodies:

$$\begin{aligned} Z &= \sup\Big\{\int \langle u, F\rangle \ :\ u \in \mathcal{D}\big(\mathbb{R}^d \setminus \Upsilon_0;\mathbb{R}^d\big),\ \ \rho\big(e(u)\big) \le 1 \ \text{ in } \overline{\Omega}\Big\} \\ &= \min\Big\{\int \rho^0(\varsigma) \ :\ \varsigma \in \mathcal{M}\big(\overline{\Omega};\mathcal{S}^{d\times d}\big),\ \ -\mathrm{Div}\,\varsigma = F \ \text{ in } \mathcal{D}'\big(\mathbb{R}^d \setminus \Upsilon_0;\mathbb{R}^d\big)\Big\}. \end{aligned}$$

Above we identify the measure F with distribution $\mathscr{F} = \mathscr{F}(u) = \int \langle u, F\rangle$ that it induces; Div is the distributional divergence operator that for matrix valued measure ς returns a vector valued distribution (essentially $\mathrm{Div} := -A^*$ for $A = \nabla$). The distributional equilibrium equation $-\mathrm{Div}\,\varsigma = F$ is in fact the *virtual work principle*:

$$\int_{\overline{\Omega}} \langle e(\phi), \varsigma\rangle = \int_{\overline{\Omega}} \langle \phi, F\rangle \qquad \forall\, \phi \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0;\mathbb{R}^d).$$

It must be stressed that above ϕ is not bound to vanish on $\partial\Omega \setminus \Upsilon_0$ hence a Neumann boundary condition is imposed on $\partial\Omega \setminus \Upsilon_0$, which can be non-homogeneous if F charges

$\partial\Omega$, cf. Example 4.34. For brevity we shall often write $-\operatorname{Div}\varsigma = F$ in $\mathscr{U}$ instead of $-\operatorname{Div}\varsigma = F$ in $\mathcal{D}'\big(\mathscr{U};\mathbb{R}^d\big)$ where $\mathscr{U}\subset\mathbb{R}^d$ is any open set.

According to Theorem 4.11 the relaxed problem $(\overline{\mathcal{P}})$ attains a solution in the set $\overline{\mathcal{U}}_1$. Below we shall give the characterization of this set for a simple scenario when $\Upsilon_0 = \varnothing$ and domain Ω has some extra regularity:

Definition 4.30. A bounded domain Ω is *star-shaped with respect to a ball* if there exists a ball $B(x_0,\varepsilon)\subset\Omega$ such that for any $x\in\Omega$ and $y\in B(x_0,\varepsilon)$ the segment $[x,y]$ is contained in Ω.

Sets that are star-shaped with respect to a ball satisfy the following dilation/contraction property: assuming that x_0 is the origin, for any $\delta\in(0,1)$ the dilated set $\Omega_\delta := (1+\delta)\,\Omega$ and the contracted set $\Omega_{-\delta} := (1-\delta)\,\Omega$ yield: $D_{\mathrm{H}}(\partial\Omega_\delta,\Omega) > 0$ and $D_{\mathrm{H}}(\partial\Omega,\Omega_{-\delta}) > 0$, where $D_{\mathrm{H}}(A,B)$ is the Hausdorff distance between sets A and B, see Section 2.3.

Proposition 4.31. *Let us assume that $\Upsilon_0 = \varnothing$ and Ω is star-shaped with respect to a ball. The characterization of the set $\overline{\mathcal{U}}_1$ follows:*

$$\overline{\mathcal{U}}_1 = \left\{ u\in\bigcap_{q=1}^{\infty} W^{1,q}(\Omega;\mathbb{R}^d)\,:\, e(u)\in L^\infty(\Omega;\mathbb{R}^d),\quad \rho\big(e(u)\big)\le 1\ \ \mathcal{L}^d\text{-a.e. in }\Omega \right\}$$

where $e(u)$ is the symmetric part of the weak gradient ∇u.

Proof. Let us first choose an element $u\in\overline{\mathcal{U}}_1$ and let $\{u_n\}\subset\mathcal{U}_1$ be a sequence furnishing uniform convergence $u_n\to u$. From Korn's inequality combined with Proposition 3.13 we obtain for any $q\in(1,\infty)$

$$\|u_n\|_{W^{1,q}(\Omega)} \le C_1\Big(\|u_n\|_{L^q(\Omega)} + \|e(u_n)\|_{L^q(\Omega)}\Big) \le C_2\Big(\|u_n\|_{L^\infty(\Omega)} + \|\rho\big(e(u_n)\big)\|_{L^\infty(\Omega)}\Big) \le C_3$$

thus, up to choosing a subsequence, u_n converges weakly in $W^{1,q}(\Omega)$ for any $q\in(1,\infty)$ while the limit must be non other than u. In addition, by the weak lower semi-continuity of L^q-norm

$$\|\rho\big(e(u)\big)\|_{L^\infty(\Omega)} = \lim_{q\to\infty}\|\rho\big(e(u)\big)\|_{L^q(\Omega)} \le \lim_{q\to\infty}\liminf_{n\to\infty}\|\rho\big(e(u_n)\big)\|_{L^q(\Omega)} \le \lim_{q\to\infty}|\Omega|^{1/q} = 1$$

establishing the "$\subset$ part" of the asserted equality.

Next we take a function $u\in W^{1,q}(\Omega;\mathbb{R}^d)$ for $q>d$ such that $\|\rho\big(e(u)\big)\|_{L^\infty(\Omega)}\le 1$. For $\Omega_\delta = (1+\delta)\,\Omega$ where $\delta>0$ we define $u_\delta = u_\delta(x) = (1+\delta)\,u\big(x/(1+\delta)\big)$ which is an element of $W^{1,q}(\Omega_\delta;\mathbb{R}^d)$ that again satisfies $\|\rho\big(e(u_\delta)\big)\|_{L^\infty(\Omega_\delta)}\le 1$. Since $u\in C(\overline{\Omega};\mathbb{R}^d)$ we obtain uniform convergence $u_\delta|_\Omega\to u$ when $\delta\to 0$. For each δ we denote $\varepsilon(\delta) = D_{\mathrm{H}}(\partial\Omega_\delta,\Omega) > 0$. Next, for $\varepsilon\in\big(0,\varepsilon(\delta)\big)$ let $u_{\delta,\varepsilon}(x) := (u_\delta * \eta_\varepsilon)(x)$ for $x\in\Omega$, where η_ε is the standard, radial-symmetric mollifying kernel satisfying $\operatorname{spt}\eta_\varepsilon = B(0,\varepsilon)$. For each δ the function $u_{\delta,\varepsilon}\in C^\infty(\overline{\Omega};\mathbb{R}^d)$ converges uniformly to $u_\delta|_\Omega$ when $\varepsilon\to 0$. Moreover, thanks to Jensen's inequality and the averaging property of mollification

$$\rho\big(e(u_{\varepsilon,\delta})\big) = \rho\left(\int_{\mathbb{R}^d} e(u_\delta)(y)\,\eta_\varepsilon(x-y)\,\mathcal{L}^d(dy)\right) \le \int_{\mathbb{R}^d}\rho\big(e(u_\delta)(y)\big)\,\eta_\varepsilon(x-y)\,\mathcal{L}^d(dy) \le 1.$$

The function $u_{\delta,\varepsilon} \in C^\infty(\overline{\Omega};\mathbb{R}^d)$ may be extended to $\overline{u}_{\delta,\varepsilon} \in \mathcal{D}(\mathbb{R}^d;\mathbb{R}^d)$ and thus $\overline{u}_{\delta,\varepsilon} \in \mathcal{U}_1$. By the diagonalization argument we may find a sequence $\overline{u}_{\delta_n,\varepsilon_n}$ that converges uniformly to u on Ω concluding the proof. □

Remark 4.32. The infinite intersection $\bigcap_{q=1}^{\infty} W^{1,q}(\Omega;\mathbb{R}^d)$ in the characterization of $\overline{\mathcal{U}}_1$ cannot be replaced by $W^{1,\infty}(\Omega;\mathbb{R}^d)$ being the space of vector valued Lipschitz continuous functions. In other words, the set $\overline{\mathcal{U}}_1$ contains functions that are not Lipschitz continuous which is reflected by the lack of Korn's inequality for exponent $q = \infty$, cf. [74].

Remark 4.33. If the space $\mathcal{U}$ is modified as follows: $\mathcal{U} = \big\{u \in \mathcal{D}(\mathbb{R}^d;\mathbb{R}^d) \,:\, u = 0 \text{ on } \Upsilon_0\big\}$ (i.e. the support of u is allowed to "touch" Υ_0) the respectively defined set $\overline{\mathcal{U}}_1$ enjoys the characterization from Proposition 4.31 if condition $u = 0$ on Υ_0 is added. A very technical proof of this fact may be found in [41, Theorem 3.1]. The present author expects that this characterization remains true for $\mathcal{U}$ defined in this thesis but ultimately this matter was not investigated here.

Owing to the characterization of $\overline{\mathcal{U}}_1$ in Proposition 4.31 we are ready to approach (FMD) analytically for a fairly general class of 2D and 3D problems – the idea is to exploit the optimality conditions. For simplicity (to avoid manipulation with constants Z and C_0), however, instead of using criteria in Corollary 4.24 we shall use criteria given in Theorem 4.22 that for a quadruple $(u,\mu,\sigma,\mathscr{H})$ verify: optimality of u and $\sigma\mu$ for problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$, respectively, and anisotropy optimality $\mathscr{H}(x) \in \underline{\mathcal{K}}_1\big(\sigma(x)\big)$ for μ-a.e. x. Then, solution λ of (FMD) will follow by Corollary 4.7. Accordingly, for: Ω star-shaped with respect to a ball, empty Dirichlet region $\Upsilon_0 = \varnothing$ and a balanced load F we must verify the following set of optimality conditions:

$$\begin{cases} \text{(i)} & -\mathrm{Div}(\sigma\mu) = F \quad \text{in } \mathbb{R}^d, \\ \text{(ii)} & u \in W^{1,q}(\Omega;\mathbb{R}^d), \quad \rho\big(e(u)\big) \le 1 \;\; \mathcal{L}^d\text{-a.e. in } \Omega, \\ \text{(iii)} & \sigma \in \partial j_\mu\big(x,\mathscr{H},e_\mu(u)\big) \quad \mu\text{-a.e.} \end{cases} \tag{4.25}$$

where $u \in C(\overline{\Omega};\mathbb{R}^d)$, $\mu \in \mathcal{M}_+(\overline{\Omega})$, $\sigma \in L^\infty_\mu(\overline{\Omega};\mathcal{S}^{d\times d})$ and $\mathscr{H} \in L^\infty_\mu(\overline{\Omega};\mathcal{K})$ with $\rho^0(\sigma) = 1$ and $\mathscr{H} \in \mathcal{K}_1$ μ-a.e.

Owing to Proposition 2.67 the tangent space $T_\mu \subset \mathbb{R}^d$ and the space of tangential matrices $M^e_\mu \subset \mathcal{S}^{d\times d}$ enjoy a simple characterization for multi-junction measures $\mu = \sum_{i=1}^N \mu_i = \sum_{i=1}^N m_i\,\mathcal{H}^{k_i} \llcorner S_i$ (cf. Definition 2.58):

$$M^e_\mu = T_\mu \otimes T_\mu = T_{S_i} \overset{s}{\otimes} T_{S_i}, \quad e_\mu(u) = P_\mu\, e(u)\, P_\mu = P_{S_i}\, e(u)\, P_{S_i} \qquad \mu_i\text{-a.e.} \tag{4.26}$$

for $u \in \mathcal{D}(\mathbb{R}^d;\mathbb{R}^d)$ (it is straightforward to show that the same formula holds for $u \in C^1(\overline{\Omega};\mathbb{R}^d)$).

The idea behind a very simple example of optimal design solved below is to compare several settings of the (FMD) for 2D elasticity. As we shall see in each case the competitor u will be smooth thus in view of Remark 4.23 we could omit the μ-tangential apparatus and check the condition $\sigma \in \partial j\big(\mathscr{H},e(u)\big)$ instead of (iii) in (4.25) above. Notwithstanding this, we shall stick to condition (iii) for illustrative purposes.

Example 4.34 (Optimal material design of a plate under uni-axial tension test). We consider a rectangle being a closed set $R = A_1A_2B_2B_1 \subset \mathbb{R}^d = \mathbb{R}^2$ (we set $d = 2$ throughout this example) with $A_1 = (-a/2, -b/2)$, $A_2 = (-a/2, b/2)$, $B_1 = (a/2, -b/2)$ and $B_2 = (a/2, b/2)$, see Figure 4.1. For Ω we can take any domain such that $R \subset \overline{\Omega}$. in particular we may choose $\Omega = \operatorname{int} R$. With an empty Dirichlet region $\Upsilon_0 = \varnothing$ we apply a load $F = F_q + F_Q$ such that

$$F_q = -q\,\mathrm{e}_1\,\mathcal{H}^1 \llcorner [A_1, A_2] + q\,\mathrm{e}_1\,\mathcal{H}^1 \llcorner [B_1, B_2], \qquad F_Q = -Q\,\mathrm{e}_1\,\delta_{A_0} + Q\,\mathrm{e}_1\,\delta_{B_0} \tag{4.27}$$

where $\mathrm{e}_1, \mathrm{e}_2$ form a Cartesian basis, $A_0 = (-a/2, 0)$, $B_0 = (a/2, 0)$ and q, Q are non-negative constants that represent, respectively, loads diffused along segments and point loads. It is straightforward to check that F is balanced.

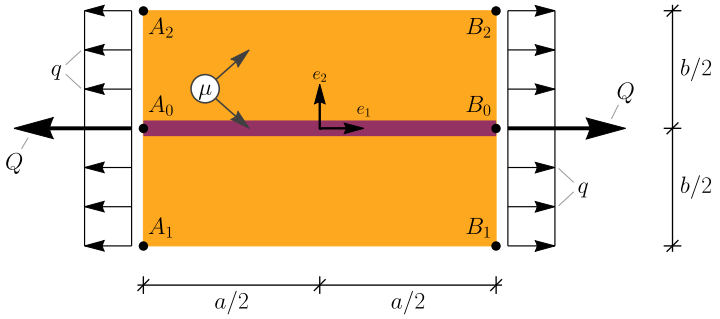


Figure 4.1: Graphical representation of load $F = F_q + F_Q$ and optimal mass distribution μ.

Case a) the Anisotropic Material Design

We first consider the AMD setting studied in Example 3.16, we recall that $\rho = \rho^0 = |\cdot|$. We propose the quadruple

$$u(x) = x_1\,\mathrm{e}_1, \qquad \mu = q\,\mathcal{L}^2 \llcorner R + Q\,\mathcal{H}^1 \llcorner [A_0, B_0],$$
$$\sigma = \mathrm{e}_1 \otimes \mathrm{e}_1, \qquad \mathscr{H} = \mathrm{e}_1 \otimes \mathrm{e}_1 \otimes \mathrm{e}_1 \otimes \mathrm{e}_1.$$

We see that $\rho^0(\sigma) = |\sigma| = 1$ and $c(\mathscr{H}) = \operatorname{Tr} \mathscr{H} = 1$, which are the initial assumptions in Theorem 4.22. An elementary use of classical integration by parts formula (in R and independently on $[A_0, B_0]$) shows that $-\operatorname{Div}(\sigma\mu) = F$, which gives optimality condition (i) in (4.25). The function u is smooth and thus checking the condition (ii) boils down to verifying whether $\rho\big(e(u)\big) = |e(u)| \leq 1$ – this is straightforward since $e(u) = \mathrm{e}_1 \otimes \mathrm{e}_1$. In order to verify condition (iii) we first list essential elements of μ-tangential calculus. The measure μ is a multi-junction measure $\mu = \mu_1 + \mu_2$ with $\mu_1 = q\,\mathcal{L}^2 \llcorner R$ and $\mu_2 = Q\,\mathcal{H}^1 \llcorner [A_0, B_0]$ hence according to (4.26) for μ-a.e. x we have:

$$M^e_\mu(x) = \begin{cases} \mathcal{S}^{2\times 2} & \text{for } \mathcal{L}^2\text{-a.e. } x \in R, \\ \operatorname{span}\{\mathrm{e}_1 \otimes \mathrm{e}_1\} & \text{for } \mathcal{H}^1\text{-a.e. } x \in [A_0, B_0], \end{cases}$$

$$P_\mu(x) = \begin{cases} \mathrm{I} & \text{for } \mathcal{L}^2\text{-a.e. } x \in R, \\ \mathrm{e}_1 \otimes \mathrm{e}_1 & \text{for } \mathcal{H}^1\text{-a.e. } x \in [A_0, B_0]. \end{cases}$$

Since u is smooth we simply compute $e_\mu(u)(x) = P_\mu(x)\,e(u)(x)\,P_\mu(x)$; having $e(u) = \mathrm{e}_1 \otimes \mathrm{e}_1$ we clearly obtain that $e_\mu(u) = \mathrm{e}_1 \otimes \mathrm{e}_1$ μ-a.e. as well. Next we compute the formula

for $j_\mu\big(x, \mathscr{H}(x), \cdot\,\big)$ using formula (4.12). For $\mathcal{L}^2$-a.e. $x \in R$ clearly $j_\mu\big(x, \mathscr{H}(x), \xi\big) = j\big(\mathscr{H}(x), \xi\big)$ since for such x we have $N^e_\mu(x) = (M^e_\mu(x))^\perp = \{0\}$. For $\mathcal{H}^1$-a.e. $x \in [A_0, B_0]$ we have $\langle \mathrm{e}_1 \otimes \mathrm{e}_1, \zeta\rangle = 0$ whenever $\zeta \in N^e_\mu(x)$ hence for any $\xi \in \mathcal{S}^{2\times 2}$

$$j_\mu\big(x, \mathscr{H}(x), \xi\big) = \inf_{\zeta \in N^e_\mu(x)} j\big(\mathscr{H}(x), \xi + \zeta\big) = \inf_{\zeta \in N^e_\mu(x)} \frac{1}{2}\big(\langle \mathrm{e}_1 \otimes \mathrm{e}_1, \xi + \zeta\rangle\big)^2 = \frac{1}{2}\big(\langle \mathrm{e}_1 \otimes \mathrm{e}_1, \xi\rangle\big)^2$$

and ultimately we obtain $j_\mu(x, \mathscr{H}, \cdot\,) = j(\mathscr{H}, \cdot\,)$ μ-a.e. Therefore, verifying condition (iii) in (4.25) boils down to verifying if μ-a.e. $\sigma \in \partial j\big(\mathscr{H}, e_\mu(u)\big)$ or, since $j(H, \xi) = \frac{1}{2}\langle H, \xi, \xi\rangle$, to $\sigma = \mathscr{H}\, e_\mu(u)$, which is straightforward to check. Eventually, the quadruple $(u, \mu, \sigma, \mathscr{H})$ is optimal in the sense of Theorem 4.22, while from Corollary 4.7 we infer that the constitutive field

$$\lambda = \frac{C_0}{Z}\, \mathscr{H}\mu, \qquad \text{where} \quad Z = \int \langle u, F\rangle = \int \rho^0(\sigma) d\mu = q\, ab + Q\, a$$

solves the (FMD) problem in the AMD setting and in particular that $\mathscr{H} \in \underline{\mathcal{K}}_1(\sigma)$ μ-a.e. (which independently follows by characterization of the set $\underline{\mathcal{K}}_1(\sigma)$, cf. (3.19) in Example 3.16).

Case b) the Fibrous Material Design

In the case of Fibrous Material Design (cf. Example 3.18) it is enough to shortly note that the quadruple $(u, \mu, \sigma, \mathscr{H})$ proposed in Case a) is also optimal in the setting of FibMD problem: indeed, both $e(u)$ and σ are of rank one, thus spectral norm $\rho\big(e(u)\big)$ and its polar $\rho^0(\sigma)$ (see (3.28)) coincide with $|e(u)|$ and $|\sigma|$ respectively. Moreover, $\mathscr{H} \in \mathcal{K}_{axial} \subset \mathcal{K}_{fib}$ while the function j in the FibMD setting is identical as in the AMD setting.

Furthermore, the same quadruple is optimal in the FibMD$\pm$ setting (cf. Example 3.32) provided that one assumes $\kappa_+ = 1$. This is a consequence of $e(u)$ and σ being positive semi-definite μ-a.e. First, $\max_i \lambda_i\big(e(u)\big) = 1 = 1/\kappa_+$ while $\min_i \lambda_i\big(e(u)\big) = 0 \geq 1/\kappa_-$ for any $\kappa_- > 0$ therefore $\rho_\pm\big(e(u)\big) = 1$ yielding optimality condition (ii). Next, for any $\xi \in \mathcal{S}^{d\times d}_+$ and any $H \in \mathcal{K}$

$$\begin{aligned} j_{\pm,\mu}\big(x, H, \xi\big) &= \inf_{\zeta \in N^e_\mu(x)} j_\pm\big(H, \xi + \zeta\big) \leq j_\pm\big(H, \xi\big) = (\kappa_+)^2 j_+\big(H, \xi\big) + (\kappa_-)^2 j_-\big(H, \xi\big) \\ &= (\kappa_+)^2 \inf_{\zeta_1 \in \mathcal{S}^{d\times d}_+} j\big(H, \xi + \zeta_1\big) \;+\; (\kappa_-)^2 \inf_{\zeta_2 \in \mathcal{S}^{d\times d}_-} j\big(H, \xi + \zeta_2\big) \\ &\leq (\kappa_+)^2 j\big(H, \xi\big) + (\kappa_-)^2 j\big(H, \xi + (-\xi)\big) = j\big(H, \xi\big) \end{aligned}$$

and, owing to (3.46), for $\sigma \in \mathcal{S}^{d\times d}_+ \cap M^e_\mu(x)$

$$j^*_{\pm,\mu}\big(x, H, \sigma\big) = j^*_\pm\big(H, \sigma\big) \leq \frac{1}{(\kappa_+)^{p'}}\, j^*_+(H, \sigma) + \frac{1}{(\kappa_-)^{p'}}\, j^*_-(H, 0) = j^*(H, \sigma).$$

Therefore, for the quadruple proposed in Case a), $j_{\pm,\mu}\big(x, \mathscr{H}, e_\mu(u)\big) + j^*_{\pm,\mu}\big(x, \mathscr{H}, \sigma\big) \leq j\big(\mathscr{H}, e_\mu(u)\big) + j^*\big(\mathscr{H}, \sigma\big) = \langle e_\mu(u), \sigma\rangle$ (the equality was established in Case a)) hence $\sigma \in \partial j_{\pm,\mu}\big(x, \mathscr{H}, e_\mu(u)\big)$, which validates optimality condition (iii).

Case c) the Isotropic Material Design

For the IMD setting, examined in Example 3.17, the norms ρ and ρ^0 are given in (3.20) and (3.21), respectively. We put forward a quadruple that shall be checked for optimality:

$$u(x) = \frac{2+\sqrt{2}}{2}\, x_1\, \mathrm{e}_1 - \frac{2-\sqrt{2}}{2}\, x_2\, \mathrm{e}_2, \qquad \mu = \frac{2+\sqrt{2}}{2}\, \Big(q\, \mathcal{L}^2 \llcorner R + Q\, \mathcal{H}^1 \llcorner [A_0, B_0] \Big),$$
$$\sigma = \frac{2}{2+\sqrt{2}}\, \mathrm{e}_1 \otimes \mathrm{e}_1, \qquad \mathscr{H} = 2\mathscr{K} \left(\frac{1}{2}\, \mathrm{I} \otimes \mathrm{I} \right) + 2\mathscr{G} \left(\mathrm{Id} - \frac{1}{2} \mathrm{I} \otimes \mathrm{I} \right)$$

where

$$\mathscr{K} = \frac{1}{2+2\sqrt{2}}, \qquad \mathscr{G} = \frac{1}{4+2\sqrt{2}}. \tag{4.28}$$

First we check that $\operatorname{Tr} \mathscr{H} = 2\mathscr{K} + 2 \cdot 2\mathscr{G} = 1$, thus $\mathscr{H} \in \mathcal{K}_1$ as assumed in Theorem 4.22. Since the absolute stress $\varsigma = \sigma\mu$ is identical to the one from Case a) condition (i) in (4.25) clearly holds. The function u is again smooth and we may compute

$$e(u) = \frac{2+\sqrt{2}}{2}\, \mathrm{e}_1 \otimes \mathrm{e}_1 - \frac{2-\sqrt{2}}{2}\, \mathrm{e}_2 \otimes \mathrm{e}_2$$

while

$$\operatorname{Tr}\big(e(u)\big) = \sqrt{2}, \quad \big|\operatorname{dev}\big(e(u)\big)\big| = \sqrt{2}, \qquad \operatorname{Tr} \sigma = \frac{2}{2+\sqrt{2}}, \quad |\operatorname{dev} \sigma| = \frac{1}{1+\sqrt{2}}$$

yielding

$$\rho\big(e(u)\big) = \max \left\{ \frac{\big|\operatorname{Tr}\big(e(u)\big)\big|}{\sqrt{2}}\,,\, \frac{\big|\operatorname{dev}\big(e(u)\big)\big|}{\sqrt{2}} \right\} = 1, \quad \rho^0(\sigma) = \frac{|\operatorname{Tr} \sigma|}{\sqrt{2}} + \sqrt{2}\, |\operatorname{dev} \sigma| = 1,$$

which validates optimality condition (ii) in (4.25); moreover $\rho^0(\sigma) = 1$ as required in Theorem 4.22.

We move on to check the last optimality condition in (4.25). Since μ above coincides with the one from Case a) (up to multiplicative constants), the formulas for M^e_μ and P_μ derived therein are also correct here. It is thus clear that $\mathcal{L}^2$-a.e. in R we have $j_\mu(x, \mathscr{H}, \cdot\,) = j(\mathscr{H}, \cdot\,)$. Meanwhile for $\mathcal{H}^1$-a.e. $x \in [A_0, B_0]$ the tensors $\zeta \in N^e_\mu(x)$ are exactly those of the form $\zeta = \mathrm{e}_2 \overset{s}{\otimes} \tau$ where $\tau \in \mathbb{R}^2$. Hence, for $\mathcal{H}^1$-a.e. $x \in [A_0, B_0]$ after performing the minimization (being non-trivial here, see below) we obtain

$$j_\mu\big(x, \mathscr{H}(x), \xi\big) = \inf_{\tau \in \mathbb{R}^2} j\big(\mathscr{H}(x), \xi + \mathrm{e}_2 \overset{s}{\otimes} \tau\big) = \frac{1}{2}\, \frac{4\mathscr{K}\mathscr{G}}{\mathscr{K}+\mathscr{G}}\, \langle \mathrm{e}_1 \otimes \mathrm{e}_1 \otimes \mathrm{e}_1 \otimes \mathrm{e}_1, \xi \otimes \xi \rangle, \tag{4.29}$$

where constant $\frac{4\mathscr{K}\mathscr{G}}{\mathscr{K}+\mathscr{G}}$ can be readily recognized as Young modulus E, cf. (3.40). For chosen ξ the minimizer $\tau = \tau_\xi$ above is exactly the one for which $\mathscr{H}(x)\,(\xi + \mathrm{e}_2 \overset{s}{\otimes} \tau_\xi) = s\, \mathrm{e}_1 \otimes \mathrm{e}_1$ for some $s \in \mathbb{R}$, i.e. the strain $\mathrm{e}_2 \overset{s}{\otimes} \tau_\xi$ normal to μ must be such that the stress generated by the Hooke law is uni-axial. The potential j_μ thus induces the well-known uni-axial constitutive law in the bar $[A_0, B_0]$.

Upon computing: $e_\mu(u) = e(u)$ $\mathcal{L}^2$-a.e. in R and $e_\mu(u) = \frac{2+\sqrt{2}}{2}\, \mathrm{e}_1 \otimes \mathrm{e}_1$ $\mathcal{H}^1$-a.e. on $[A_0, B_0]$, we see that verifying condition (iii) boils down to checking if

$$\sigma(x) = \begin{cases} \mathscr{H}(x)\, e(u)(x) & \text{for } \mathcal{L}^2\text{-a.e. } x \in R, \\ \big(\frac{4\mathscr{K}\mathscr{G}}{\mathscr{K}+\mathscr{G}}\, \mathrm{e}_1 \otimes \mathrm{e}_1 \otimes \mathrm{e}_1 \otimes \mathrm{e}_1\big)\, \big(\frac{2+\sqrt{2}}{2} \mathrm{e}_1 \otimes \mathrm{e}_1\big) & \text{for } \mathcal{H}^1\text{-a.e. } x \in [A_0, B_0]. \end{cases}$$

The equations above may be verified through elementary computations. We readily see that the quadruple $(u, \mu, \sigma, \mathscr{H})$ is optimal in the sense of Theorem 4.22 and the optimal Hooke tensor field follows by Corollary 4.7:

$$\lambda = \frac{C_0}{Z}\,\mathscr{H}\mu, \qquad \text{where} \quad Z = \int \langle u, F\rangle = \int \rho^0(\sigma)d\mu = \frac{2+\sqrt{2}}{2}\big(q\,ab + Q\,a\big).$$

We conclude the example by computing the Young modulus and the Poisson ratio corresponding to optimal $\mathscr{K}, \mathscr{G}$ via formulas (3.40)

$$E = \frac{4\mathscr{K}\mathscr{G}}{\mathscr{K}+\mathscr{G}} = \left(\frac{2}{2+\sqrt{2}}\right)^2 = 6 - 4\sqrt{2}, \qquad \nu = \frac{\mathscr{K}-\mathscr{G}}{\mathscr{K}+\mathscr{G}} = 3 - 2\sqrt{2}. \tag{4.30}$$

The computed value of Young modulus E turns out to be maximal among all pairs $\mathscr{K}, \mathscr{G} \geq 0$ satisfying $\operatorname{Tr}\mathscr{H} = 2\mathscr{K} + 2\cdot 2\mathscr{G} \leq 1$. △

4.5.2. Framework #2: Free Material Design of Kirchhoff plates

We present the framework of Kirchhoff plates where $\Omega \subset \mathbb{R}^d = \mathbb{R}^2$ is a planar bounded design domain (we shall keep the dimension symbol d nonetheless). Assuming that the boundary $\partial\Omega$ is Lipschitz continuous we pose the (FMD) problem for the following data:

$$\begin{cases} U = \mathbb{R}, \quad \mathcal{U} = \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0), \\ |||u||| = \|u\|_{\infty,\Omega} + \|\nabla u\|_{\infty,\Omega}, \\ \mathscr{F}(u) = \int u\,df + \int \langle \nabla u, \mathscr{M}\rangle \quad \text{where } f \in \mathcal{M}(\overline{\Omega};\mathbb{R}),\ \mathscr{M} \in \mathcal{M}(\overline{\Omega};\mathbb{R}^d), \\ V = \mathcal{S}^{d\times d}, \quad Au = \nabla^2 u := \nabla\big(\nabla u\big), \\ j : \mathcal{K}\times\mathcal{S}^{d\times d} \to \mathbb{R}_+ \quad \text{satisfying (A1)-(A4) and } (\mathrm{A}_{\mathrm{ellip}}). \end{cases} \tag{4.31}$$

The function u plays the role of the plate's deflection; the signed measure f is the downward/upward load or *the pressure load* while the vector measure $\mathscr{M}$ is the moment loading, e.g. applied to the boundary if $\mathscr{M} = m\,\mathcal{H}^1 \llcorner \partial\Omega$ for $m \in L^1_{\mathcal{H}^1}(\partial\Omega;\mathbb{R}^d)$ or concentrated at a point if $\mathscr{M} = M\,\delta_{x_0}$ for $x_0 \in \overline{\Omega}$. The tensor function $Au = \nabla^2 u$ is the infinitesimal curvature of the deformed plate (classically, however, we take $-\nabla^2 u$ to fit the mechanical interpretation). In case of Kirchhoff plates the constitutive field $\lambda \in \mathcal{M}(\overline{\Omega};\mathcal{K})$ appears in (FMD) as the field of *the plate rigidity*, i.e. "point-wisely" λ should be considered a moment (with respect to the central plane of the plate) of 3D Hooke tensor field computed over the thickness of the plate. Considering the choice of the norm $|||\cdot|||$ it is easy to observe that $\overline{\mathcal{U}}_1 \subset \overline{\mathcal{U}} \subset \big\{u \in C^1(\overline{\Omega};\mathbb{R}) \,:\, u = 0,\ \nabla u = 0 \text{ on } \Upsilon_0\big\}$, namely the designed plate is assumed to be *clamped* on Υ_0.

Recalling that $\mathcal{R} = \big\{u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0) \,:\, \nabla^2 u = 0 \text{ in } \Omega\big\}$ we assume that one of the scenarios below occurs:

(1) Υ_0 is an empty set and the pair of loads $f, \mathscr{M}$ is balanced, i.e.

$$\int_{\overline{\Omega}} f = 0 \in \mathbb{R}, \qquad \int_{\overline{\Omega}} x\,f + \int_{\overline{\Omega}} \mathscr{M} = 0 \in \mathbb{R}^d. \tag{4.32}$$

In this case we choose the linear operator $P_{\mathcal{R}} : \mathcal{U} \to \mathcal{R}$ that satisfies

$$(P_{\mathcal{R}}u)(x) = a_0(u) + \langle b_0(u), x\rangle$$

where $b_0(u) = \fint_\Omega \nabla u \, d\mathcal{L}^d$ and $a_0(u) = \fint_\Omega \big(u(x) - \langle b_0(u), x\rangle\big) \, \mathcal{L}^d(dx)$

(2) or the set Υ_0 is non-empty and the loads $f, \mathscr{M}$ are arbitrary. In this case $\mathcal{R} = \{0\}$ and thus we set

$$P_{\mathcal{R}} u = 0.$$

We define a space of affine functions over Ω:

$$\mathcal{R}_\Omega = \Big\{ u = u(x) = a_0 + \langle b_0, x\rangle \ : \ a_0 \in \mathbb{R}, \ b_0 \in \mathbb{R}^d, \ u = 0, \ \nabla u = 0 \text{ on } \Upsilon_0 \Big\}$$

i.e. if deflection function is an element of $\mathcal{R}_\Omega$ then it is a vertical translation by a_0 composed with an infinitesimal rotation of the plate about an in-plane axis represented by its normal vector b_0. In addition both u and ∇u must be zero on Υ_0 thus it is easily seen that $\mathcal{R}_\Omega = \{0\}$ whenever $\Upsilon_0 \neq \varnothing$. We see that the space $\mathcal{R}_\Omega$ is the restriction of $\mathcal{R}$ to Ω. Moreover in both cases (1) and (2) $P_{\mathcal{R}}$ are projections being linear continuous operators from $W^{1,q}$ to $\mathcal{R}_\Omega$. On top of that, it holds that $\mathcal{R}_\Omega = \big\{ u \in W^{2,q}(\Omega) \ : \ \nabla^2 u = 0 \text{ in } \Omega, \ u = 0, \ \nabla u = 0 \text{ on } \Upsilon_0 \big\}$.

Remark 4.35. Unlike in Framework #1 we do not make any assumptions on the set Υ_0, cf. Remark 4.28. This is due to the fact that here $\mathcal{R}_\Omega$ (the correct space of rigid body motions) is always the restriction of $\mathcal{R}$ to Ω. However, we stumble on a limitation of other nature: within theory herein presented we cannot design plates that are e.g. *simply supported* on the boundary $\partial\Omega$. For that purpose one would have to define two sets $\Upsilon_0, \Upsilon_1 \subset \Upsilon_0$ and then the space of admissible smooth deflections $\mathcal{U} = \big\{ u \in \mathcal{D}(\mathbb{R}^d) : u = 0 \text{ on } \Upsilon_0, \ \nabla u = 0 \text{ on } \Upsilon_1 \big\}$. This idea, however, lies beyond the scope of this thesis and we shall limit our considerations to plates that are clamped on Υ_0.

We may now verify assumptions (B1)-(B5). The continuity condition (B2) is straightforward considering the choice of $\|\!|\cdot|\!\|$ while (B1) may be checked easily in both cases (1), (2) above using the fact that $\mathcal{R}_\Omega$ is the restriction of $\mathcal{R}$ to Ω. The μ-tangential calculus for operator $A = \nabla^2$ has been already studied in Section 2.5.4 where assumption (B4) was verified (condition ($\mathrm{H}_{\mathrm{stab}}$) therein). Assumption (B5) or, equivalently, condition ($\mathrm{A}_{\mathrm{ellip}}$) is explicitly given in (4.31). Finally, the compactness assumption (B3) is due to the following result (that was already proved in [37] for $\Upsilon_0 = \varnothing$):

Proposition 4.36. *Assuming that the boundary $\partial\Omega$ is Lipschitz continuous and* ($\mathrm{A}_{\mathrm{ellip}}$) *holds, for any $C > 0$ the set*

$$\Big\{ u - P_{\mathcal{R}} u \ : \ u \in \mathcal{D}(\mathbb{R}^d), \ \|\rho\big(\nabla^2 u\big)\|_{\infty,\Omega} \leq C \Big\} \tag{4.33}$$

is pre-compact in $C^1(\overline{\Omega}; \mathbb{R}^d)$.

Proof. Let us first assume that $\Upsilon_0 = \varnothing$ and set $\tilde{u} = u - P_{\mathcal{R}} u$. Then $\fint \tilde{u} \, d\mathcal{L}^d = 0$, $\fint \nabla \tilde{u} \, d\mathcal{L}^d = 0$ and for any $q \in [1, \infty)$ by using Poincaré inequality twice we obtain

$$\|\tilde{u}\|_{L^q(\Omega)} \leq C_1 \|\nabla \tilde{u}\|_{L^q(\Omega)} \leq C_2 \|\nabla^2 \tilde{u}\|_{L^q(\Omega)} = C_3 \|\nabla^2 u\|_{L^q(\Omega)} \leq C_4 \|\rho\big(\nabla^2 u\|\big)_{L^\infty(\Omega)}$$

where in the last inequality we have used assumption ($\mathrm{A}_{\mathrm{ellip}}$), cf. Proposition 3.13. The set (4.33) is thus bounded in $W^{2,q}(\Omega)$ for any $q \in [1, \infty)$ and the claim follows by Morrey's embedding theorem for $q > d$.

In the case when $\Upsilon_0 \neq \varnothing$ we have $\mathcal{R} = \{0\}$ and thus $P_{\mathcal{R}}u = 0$. To prove that there exist constants C_1 and C_2 such that $\|u\|_{L^q(\Omega)} \leq C_1\|\nabla u\|_{L^q(\Omega)}$ and, independently, $\|\nabla u\|_{L^q(\Omega)} \leq C_2\|\nabla^2 u\|_{L^q(\Omega)}$ for any $u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0)$ we argue by contradiction similarly as in the proof of Proposition 4.29 (we do not require Korn's inequality though). The rest of the proof remains the same as for $\Upsilon_0 = \varnothing$. □

All assumptions (B1)-(B5) are therefore satisfied and the theory of (FMD) problem is ready to use for Kirchhoff plates; in particular the (FMD) problem always attains a solution $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$. Below we give the specification of problems $(\mathcal{P})$ and $(\mathcal{P}^*)$ for Kirchhoff plates:

$$\begin{aligned} Z &= \sup\Big\{\int u\,df + \int \langle \nabla u, \mathscr{M}\rangle \ : \ u \in \mathcal{D}\big(\mathbb{R}^d \setminus \Upsilon_0\big), \ \ \rho\big(\nabla^2 u\big) \leq 1 \ \text{ in } \overline{\Omega}\Big\} \\ &= \min\Big\{\int \rho^0(\varsigma) \ : \ \varsigma \in \mathcal{M}\big(\overline{\Omega}; \mathcal{S}^{d\times d}\big), \ \ \operatorname{div}^2\varsigma = f - \operatorname{Div}\mathscr{M} \ \text{ in } \mathcal{D}'\big(\mathbb{R}^d \setminus \Upsilon_0\big)\Big\}. \end{aligned}$$

Above the tensor valued measure $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d})$ represents a (absolute) bending moment field. Similarly to $\operatorname{Div} = -(\nabla)^*$, the double divergence operator div^2 is intended in the distributional sense, i.e. $\operatorname{div}^2 := A^*$ for $A = \nabla^2$ (by different symbols div and Div we distinguish between operators that produce real and, respectively, vector valued distributions).

In order to effectively solve the (FMD) problem we require a characterization of the set $\overline{\mathcal{U}}_1$, which was derived in [37] for the case $\Upsilon_0 = \varnothing$. Its generalization to arbitrary Υ_0 is very technical hence we will settle for a simpler result where $\Upsilon_0 = \partial\Omega$, i.e. the plate is fully clamped on the boundary:

Proposition 4.37. *Let us assume that Ω is star-shaped with respect to a ball. Then:*

- *for $\Upsilon_0 = \varnothing$*
$$\overline{\mathcal{U}}_1 = \Big\{u \in W^{2,\infty}(\Omega;\mathbb{R}) \ : \ \rho\big(\nabla^2 u\big) \leq 1 \ \ \mathcal{L}^d\text{-a.e. in } \Omega\Big\};$$
- *for $\Upsilon_0 = \partial\Omega$*
$$\overline{\mathcal{U}}_1 = \Big\{u \in W^{2,\infty}(\Omega;\mathbb{R}) \ : \ u = 0,\, \nabla u = 0 \text{ on } \partial\Omega, \ \ \rho\big(\nabla^2 u\big) \leq 1 \ \ \mathcal{L}^d\text{-a.e. in } \Omega\Big\}.$$

Proof. In case of $\Upsilon_0 = \varnothing$ the proof is analogous to the proof of Proposition 4.31 up to acknowledging that $\|\rho(\nabla^2 u)\|_{L^\infty(\Omega)} \leq 1$ implies that $u \in W^{2,\infty}(\Omega)$.

When $\Upsilon_0 = \partial\Omega$ the proof of inclusion $\subset$ is again similar as for $\Upsilon_0 = \varnothing$: for each $u \in \mathcal{U}_1$ there holds $u|_\Omega \in \mathcal{D}(\Omega)$ thus each cluster point of $\mathcal{U}_1$ in $C^1(\overline{\Omega};\mathbb{R})$ topology must yield $u = 0, \nabla u = 0$ on $\partial\Omega$.

To prove the converse inclusion for $\Upsilon_0 = \partial\Omega$ we take $u \in W^{2,\infty}(\Omega)$ with $u = 0, \nabla u = 0$ on $\partial\Omega$. Then, the extension $\overline{u}$ of u to whole $\mathbb{R}^d$ by zero is an element of $W^{2,\infty}(\mathbb{R}^d)$ satisfying $\rho(\nabla^2\overline{u})$ a.e. for $\delta > 0$ we define the contraction $\overline{u}_\delta = \overline{u}_\delta(x) = 1/(1+\delta)^2\,\overline{u}\big((1+\delta)\,x\big)$ (as opposed to dilation in case of $\Upsilon_0 = \varnothing$). Then it is easy to check that again $\rho(\nabla^2\overline{u}_\delta) \leq 1$ a.e. while $D_{\mathrm{H}}\big(\partial\Omega, \operatorname{spt}\overline{u}_\delta\big) = \varepsilon(\delta) > 0$ thus $\overline{u}_{\delta,\varepsilon} = \overline{u}_\delta * \eta_\varepsilon$ is an element of $\mathcal{D}(\Omega)$ for $\varepsilon \leq \varepsilon(\delta)$. Similarly as in the proof of Proposition 4.31 inequality $\rho(\overline{u}_{\delta,\varepsilon}) \leq 1$ follows by Jensen's

inequality while sequence $\overline{u}_{\delta_n,\varepsilon_n}$ that is convergent to u in $C^1(\overline{\Omega};\mathbb{R})$ may readily be found by diagonalization argument. □

Remark 4.38. The characterization result for $\Upsilon_0 = \partial\Omega$ could be obtained just as easily in the Framework #1. Contrarily to Kirchhoff plates, however, the problem of 2D or 3D body that is fixed on the whole boundary $\partial\Omega$ is rather unnatural.

Readily, we can write down the optimality conditions from Theorem 4.22 in the form specific to Kirchhoff plates; we concentrate on the case when $\Upsilon_0 = \varnothing$. A quadruple $(u, \mu, \sigma, \mathscr{H})$ is a solution in the sense of Theorem 4.22 if and only if the following optimality criteria are met:

$$\begin{cases} \text{(i)} & \operatorname{div}^2 \varsigma = f - \operatorname{Div} \mathscr{M} \ \text{ in } \mathcal{D}'\big(\mathbb{R}^d\big), \\ \text{(ii)} & u \in W^{2,\infty}(\Omega;\mathbb{R}), \quad \rho\big(\nabla^2 u\big) \le 1 \ \ \mathcal{L}^d\text{-a.e. in } \Omega, \\ \text{(iii)} & \sigma \in \partial j_\mu\big(x, \mathscr{H}, \nabla^2_\mu u\big) \quad \mu\text{-a.e.} \end{cases} \tag{4.34}$$

where in addition we enforce that $\rho^0(\sigma) = 1$ and $\mathscr{H} \in \mathcal{K}_1 \ \ \mu$-a.e.

Drawing upon Section 2.5.4 we recall the basic elements of the μ-tangential calculus for the Hessian operator $A = \nabla^2$. For a multi-junction measure $\mu = \sum_{i=1}^N \mu_i$ we obtain

$$\mathcal{S}^{d\times d} = M_\mu^{\nabla^2} \oplus \Big(T_{S_i}^\perp \overset{s}{\otimes} T_{S_i}^\perp\Big), \quad \nabla^2_\mu u = P_{S_i}\,\nabla^2 u\,P_{S_i} + P_{S_i}^\perp\,\nabla^2 u\,P_{S_i} + P_{S_i}\,\nabla^2 u\,P_{S_i}^\perp \qquad \mu_i\text{-a.e.} \tag{4.35}$$

where u is any smooth function (at least of C^2 class).

The Fibrous Material Design setting of the problem (FMD) for Kirchhoff plates constitutes a second-order counterpart of the Michell problem: *optimal grillage problem*, see [141]. Analysis of this optimal design problem is scheduled for the future research by the present author. In this thesis, however, we stop the study of the Free Material Design for $A = \nabla^2$ here.

4.5.3. Framework #3: Free Material Design of heat conductor

The last framework of the (FMD) problem considered in this thesis concerns optimal design of conductor subject to a stationary heat flow. For a bounded domain $\Omega \subset \mathbb{R}^d$ ($d = 2$ or $d = 3$) of Lipschitz continuous boundary we set:

$$\begin{cases} U = \mathbb{R}, \quad \mathcal{U} = \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0), \\ |||u||| = \|u\|_{\infty,\Omega}, \\ \mathscr{F}(u) = \int u\,df \quad \text{where } f \in \mathcal{M}(\overline{\Omega};\mathbb{R}), \\ V = \mathbb{R}^d, \quad Au = \nabla u, \\ j : \mathcal{K} \times \mathbb{R}^d \to \mathbb{R}_+ \quad \text{satisfying (A1)-(A4) and } (\mathrm{A}_{\mathrm{ellip}}). \end{cases} \tag{4.36}$$

The scalar function u plays a role of temperature field. Assuming that $f = f_+ - f_-$ is the uniquely determined Jordan decomposition of a signed measure f into positive measures $f_+, f_- \in \mathcal{M}_+(\overline{\Omega})$ (see (2.4)) the measure f_+ can be viewed as a heat source and f_- as a heat outlet (or the heat sink). The design domain Ω is perfectly isolated on $\partial\Omega \setminus \Upsilon_0$ while

on the set Υ_0 the temperature is fixed allowing free heat transfer. The constitutive field $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ is now a distribution of conductivity tensor and compliance is defined by:

$$\mathcal{C}(\lambda) = \sup \left\{ \int u \, df - \int j\big(\lambda, \nabla u\big) \ : \ u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0) \right\}. \tag{4.37}$$

We shall assume that one of the scenarios below occurs:

(1) Υ_0 is an empty set and f is of zero mean value:

$$\int_{\overline{\Omega}} f = 0 \in \mathbb{R}, \tag{4.38}$$

or, equivalently, $f_+(\overline{\Omega}) = f_-(\overline{\Omega})$. In this case $P_{\mathcal{R}} : \mathcal{U} \to \mathcal{R}$ will be simply the mean value operator:

$$P_{\mathcal{R}} u = \fint_{\Omega} u \, d\mathcal{L}^d$$

(2) or the set Υ_0 is non-empty and the source f is arbitrary. In this case $\mathcal{R} = \{0\}$ and thus we set

$$P_{\mathcal{R}} u = 0.$$

By arguing similarly as for the first two frameworks in Sections 4.5.1 and 4.5.2 we verify assumptions (B1)-(B5) while arriving at analogous results:

Proposition 4.39. *Assuming that the boundary $\partial\Omega$ is Lipschitz continuous and* ($\mathrm{A}_{\mathrm{ellip}}$) *holds, for any $C > 0$ the set*

$$\Big\{ u - P_{\mathcal{R}} u \ : \ u \in \mathcal{D}(\mathbb{R}^d), \ \|\rho\big(\nabla u\big)\|_{\infty,\Omega} \leq C \Big\} \tag{4.39}$$

is pre-compact in $C(\overline{\Omega}; \mathbb{R}^d)$.

Proposition 4.40. *Let us assume that Ω is star-shaped with respect to a ball. Then*

- *for $\Upsilon_0 = \varnothing$*

$$\overline{\mathcal{U}}_1 = \Big\{ u \in W^{1,\infty}(\Omega; \mathbb{R}) \ : \ \rho\big(\nabla u\big) \leq 1 \ \ \mathcal{L}^d\text{-a.e. in } \Omega \Big\},$$

- *for $\Upsilon_0 = \partial\Omega$*

$$\overline{\mathcal{U}}_1 = \Big\{ u \in W^{1,\infty}(\Omega; \mathbb{R}) \ : \ u = 0 \text{ on } \partial\Omega, \ \ \rho\big(\nabla u\big) \leq 1 \ \ \mathcal{L}^d\text{-a.e. in } \Omega \Big\}.$$

The pair of problems ($\mathcal{P}$) and ($\mathcal{P}^*$) specified for Free Material Design of heat conductor reads as below:

$$Z = \sup \Big\{ \int u \, df \ : \ u \in \mathcal{D}\big(\mathbb{R}^d \setminus \Upsilon_0\big), \ \ \rho\big(\nabla u\big) \leq 1 \ \text{ in } \overline{\Omega} \Big\} \tag{4.40}$$

$$= \min \Big\{ \int \rho^0(\vartheta) \ : \ \vartheta \in \mathcal{M}\big(\overline{\Omega}; \mathbb{R}^d\big), \ \ -\mathrm{div}\, \vartheta = f \ \text{ in } \mathbb{R}^d \setminus \Upsilon_0 \Big\}. \tag{4.41}$$

where again the equation $-\mathrm{div}\, \vartheta = f$ in $\mathbb{R}^d \setminus \Upsilon_0$ is intended in the sense of distributions $\mathcal{D}'\big(\mathbb{R}^d \setminus \Upsilon_0\big)$.

Remark 4.41. We have changed the notation: from ς, that in this thesis will be mainly reserved for stress fields, to ϑ being a vector valued measure in $\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$ that represents the heat flux. Similarly we shall write q instead of σ. Speaking more accurately from the point of view of physics, it is the measure $\widetilde{\vartheta} := -\vartheta$ that is the heat flux. Indeed, for $f = \delta_y - \delta_x$ measure $\vartheta = \frac{y-x}{|y-x|}\,\mathcal{H}^1 \llcorner [x,y]$ yields $-\operatorname{div}\vartheta = f$ while $\frac{y-x}{|y-x|}$ is directed from x to y. This is due to the fact that in the Fourier's law the gradient of temperature enters with a minus and thus one should take $A = -\nabla u$ to comply with the physical context, which would render $A^*\widetilde{\vartheta} = f$ the correct law of energy conservation: $\operatorname{div}\widetilde{\vartheta} = f$.

Assuming that $\Upsilon_0 = \varnothing$ we give the optimality conditions of the form specific to Free Material Design of heat conductor: a quadruple $(u,\mu,q,\mathscr{H})$ is a solution in the sense of Theorem 4.22 if and only if

$$\begin{cases} \text{(i)} & -\operatorname{div}\vartheta = f \ \text{ in } \mathbb{R}^d, \\ \text{(ii)} & u \in W^{1,\infty}(\Omega;\mathbb{R}), \quad \rho\big(\nabla u\big) \le 1 \ \ \mathcal{L}^d\text{-a.e. in } \Omega, \\ \text{(iii)} & q \in \partial j_\mu\big(x,\mathscr{H},\nabla_\mu u\big) \quad \mu\text{-a.e.} \end{cases} \tag{4.42}$$

where in addition there must hold $\rho^0(q) = 1$ and $\mathscr{H} \in \mathcal{K}_1 \ \ \mu$-a.e.

Assuming a multi-junction measure $\mu = \sum_{i=1}^N \mu_i$ we recall the form of the tangent space and the μ-tangential gradient:

$$T_\mu = T_{S_i}, \qquad \nabla_\mu u = P_{S_i}\nabla u \qquad \mu_i\text{-a.e.}$$

where u is any smooth function (at least of C^1 class), cf. Section 2.5.2.

Unlike in elasticity, the only natural setting of Free Material Design of heat conductor that truly involves design of anisotropy is the Anisotropic Material Design setting (AMD) (cf. Example 3.16) in which we choose

$$\mathcal{K} = \mathcal{K}_{ani,c} = \mathscr{L}_+(\mathbb{R}^d) \equiv \mathcal{S}_+^{d\times d}, \qquad j(H,\xi) = \frac{1}{2}\,\langle H\xi,\xi\rangle, \qquad c(H) = \operatorname{Tr} H. \tag{4.43}$$

where $\mathcal{K}_{ani,c}$ stands for the set of all anisotropic conductivity tensors, being isomorphic to the closed convex cone of positive semi-definite symmetric matrices $\mathcal{S}_+^{d\times d}$. In Example 3.16 we have characterized the functions and sets that define the problems $(\mathcal{P})$ and $(\mathcal{P}^*)$: for any $\xi, q \in \mathbb{R}^d$ we have obtained that

$$\rho(\xi) = |\xi|, \qquad \rho^0(q) = |q|, \qquad \bar{\mathcal{K}}_1(\xi) = \left\{ \frac{\xi}{|\xi|} \otimes \frac{\xi}{|\xi|} \right\}, \qquad \underline{\mathcal{K}}_1(q) = \left\{ \frac{q}{|q|} \otimes \frac{q}{|q|}, \right\}.$$

By applying the above to Theorem 4.6 we find that the constitutive field $\lambda \in \mathcal{M}(\overline{\Omega};\mathcal{S}_+^{d\times d})$ is a solution of (FMD) problem if and only if it is of the form:

$$\lambda = \frac{C_0}{Z}\left(\frac{d\vartheta}{d|\vartheta|} \otimes \frac{d\vartheta}{d|\vartheta|}\right)|\vartheta| \tag{4.44}$$

where ϑ is any solution of problem $(\mathcal{P}^*)$ specified in (4.41) with $\rho^0 = |\cdot|$. We conclude the following rank-one property of optimal conductors:

Proposition 4.42. *Let λ be a solution of the* (FMD) *problem for heat conductor in the* AMD *setting* (4.43)*. Then λ is rank-one, namely $\frac{d\lambda}{d|\lambda|}$ is a rank-one matrix $|\lambda|$-a.e.*

Remark 4.43. We observe that, up to a multiplicative constant, the only isotropic gauge function ρ on $\mathbb{R}^d$ is the Euclidean norm $|\cdot|$. We thus find that every "isotropic" (FMD) problem for heat conductor reduces to the pair $(\mathcal{P})$, $(\mathcal{P}^*)$ with $\rho = b\,|\cdot|$, $b > 0$. Note that no similar conclusion was true for the (FMD) problem for elastic bodies, cf. Examples 3.18, 3.32, 3.17 where all the functions ρ are isotropic. For instance in (4.43) we could instead take the set of all isotropic conductivity tensors $\mathcal{K} = \mathcal{K}_{iso,c}$ while of course: $H \in \mathcal{K}_{iso}$ if and only if $H = a\,\mathrm{I}$, for $a \geq 0$, cf. Example 3.2. Then the (FMD) problem is equivalent to the Mass Optimization Problem in [35] and, since $c(\mathrm{I}) = \operatorname{Tr}\mathrm{I} = d$, for any $\xi \in \mathbb{R}^d$ we have

$$\frac{1}{2}\big(\rho(\xi)\big)^2 = \max_{\substack{H\in\mathcal{K}_{iso}\\ c(H)\leq 1}} j(H,\xi) = \max_{\substack{a\geq 0\\ d\cdot a\leq 1}} \frac{1}{2}\langle a\,\mathrm{I}, \xi\otimes\xi\rangle = \frac{1}{d}\,\frac{1}{2}|\xi|^2$$

yielding $\rho(v) = \frac{1}{\sqrt{d}}|\xi|$ and $\rho^0(q) = \sqrt{d}\,|q|$. In dimension $d = 2$, once $\Upsilon_0 = \varnothing$ and f charges the boundary $\partial\Omega$ only, the problem $(\mathcal{P}^*)$ can be reformulated as the Least Gradient Problem, see [97]. Upon acknowledging this equivalence, a study of $(\mathcal{P}^*)$ for anisotropic functions ρ^0 can be found in [96].

With the above remark in mind we may treat the AMD setting of Free Material Design of heat conductor as the representative one. Similarly as discovered in [35] for the Mass Optimization Problem we will find that such (FMD) problem of heat conductor is equivalent to the Monge-Kantorovich optimal transport problem. This will be the topic of a separate Section 4.8.

4.6. A finite element approximation of the pair $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$

With the Free Material Design problem reduced to $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$ we may now address the issue of designing a suitable numerical method while concentrating on the latter pair of mutually dual convex problems. The point of departure will be the finite element (FE) approximation $(\mathcal{P}_h)$ of $(\overline{\mathcal{P}})$. The proposed finite element method (FEM) will be tailored for the first-order frameworks of (FMD) problem, i.e. for the elasticity Framework #1 where $A = e$ or the stationary heat conduction Framework #3 where $A = \nabla$. It allows to obtain constant values of Au on each finite element. Then the key idea consists in rewriting $(\mathcal{P}_h)$ as a finite dimensional conic programming problem (cf. Section 2.6) which makes it possible to tackle problems for very fine meshes. Although the algorithm is not limited to any dimension of the design space only the 2D version will be discussed in details and then implemented.

Upfront it must be stressed that convergence of the FEM (and of other numerical methods in this dissertation) lies beyond the scope of this thesis and will not be studied. The purpose of the method is different and it is twofold. First, it is to deliver high-resolution simulations of optimal designs: it will allow to discern how the optimal shape and material distribution changes for different anisotropy settings AMD, IMD, FibMD, FibMD$_\pm$, MMD. Secondly, the present author's intention is to show that, aside from the theoretical/analytical benefits, the pair $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$ has truly practical implications. It opens up possibilities for efficient numerical methods that are alternative to the ones usually used in the literature

where (FMD) problem is tackled directly, see e.g. [168, 169], [110, 102]. It is important to note that the finite element approximation of the pair $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$ for AMD and IMD settings has been already developed in [64, 62, 68]. The formulation in the latter papers, however, is very different from the one that is to be presented below.

4.6.1. Towards a conic programming problem

For simplicity we shall assume that the design domain Ω is polyhedral (or polygonal for $d = 2$). For a mesh parameter $h > 0$ let there be given a triangulation being a family of $n = n(h)$ open disjoint simplices of diameters at most equal to h:

$$\mathcal{T}^h = \Big\{ E_i^h \,:\, E_i^h \text{ is an open simplex, } E_{i_1}^h \cap E_{i_2}^h = \varnothing \text{ for } i_1 \neq i_2, \ \operatorname{diam}(E_i^h) \leq h \Big\}$$

while we assume that $\overline{\Omega} = \bigcup_{i=1}^n \overline{E_i^h}$. For $\mathcal{U}^h(\Omega)$ we choose the finite dimensional space of displacement functions u which are element-wise affine:

$$\mathcal{U}^h(\Omega) := \Big\{ u \in W^{1,\infty}(\Omega;\mathbb{R}^d) \,:\, u = 0 \text{ on } \Upsilon_0, \\ u(x) = a_i + A_i x \ \ \forall\, x \in E_i^h, \ a_i \in \mathbb{R}^d, \ A_i \in \mathbb{R}^{d\times d} \Big\}.$$

The finite element approximation of the problem $(\overline{\mathcal{P}})$ may readily be given:

$$Z_h = \sup \left\{ \int_{\overline{\Omega}} \langle u, F\rangle \,:\, u \in \mathcal{U}^h(\Omega), \quad \rho\big(e(u)\big) \leq 1 \text{ a.e. in } \Omega \right\} \tag{$\mathcal{P}_h$}$$

If there are no Dirichlet conditions present, i.e. $\Upsilon_0 = \varnothing$, from characterization of the set $\overline{\mathcal{U}}_1$ in Proposition 4.31 we can see that the set of admissible functions in $(\mathcal{P}_h)$ is contained in $\overline{\mathcal{U}}_1$ and thus we immediately infer that:

Proposition 4.44. *For Ω being polyhedral let us consider triangulation $\mathcal{T}^h$. Then, assuming that $\Upsilon_0 = \varnothing$*

$$Z_h \leq Z.$$

Remark 4.45. The author expects that the result above can be extended to any Dirichlet region Υ_0 – for this purpose the characterization in Proposition 4.31 must be generalized which requires technical treatment.

We introduce the set of stress functions that satisfy the *approximate equilibrium equations*:

$$\Sigma^h(\Omega, F) := \left\{ \sigma \in L^1(\Omega;\mathcal{S}^{d\times d}) \,:\, \int_\Omega \langle e(\phi), \sigma\rangle \, d\mathcal{L}^d = \int_{\overline{\Omega}} \langle \phi, F\rangle \quad \forall \phi \in \mathcal{U}^h(\Omega) \right\}.$$

By standard duality arguments (cf. [80]) we arrive at the problem that is dual to $(\mathcal{P}_h)$:

$$\inf \left\{ \int_\Omega \rho^0(\sigma)\, d\mathcal{L}^d \,:\, \sigma \in \Sigma^h(\Omega, F) \right\} \tag{$\mathcal{P}_h^*$}$$

The problem $(\mathcal{P}_h^*)$ above is *a priori* infinite dimensional and using standard tools we merely obtain inequality $Z_h = \sup \mathcal{P}_h \leq \inf \mathcal{P}_h^*$. In due course, however, we will prove that the converse inequality holds as well while $(\mathcal{P}_h^*)$ will appear to be equivalent to a finite dimensional problem, cf. Corollary 4.49 below.

Naturally, in order to make the finite element method practical we shall parameterize the space $\mathcal{U}^h(\Omega)$ by identifying its elements with the vector values at nodes, i.e. at the vertices of triangulation $\mathcal{T}^h$. Our goal is to write down an explicit finite programming problem that is ready to be implemented. For this reason we shall fix the dimension of the design space: henceforward $d = 2$ and $\mathrm{e}_1, \mathrm{e}_2$ will be a Cartesian basis of $\mathbb{R}^2$. Generalization of the forthcoming results towards $d = 3$ may be considered straightforward. Below we will employ bold font to denote vectors that gather unknowns and data for elements and nodes, as well as matrices that establish linear relations between them. For a vector $\mathbf{v} \in \mathbb{R}^N$ by $\mathbf{v}(i)$ we shall denote its i-th component.

Let X^h be the finite set of all vertices of triangulation $\mathcal{T}^h$. We will henceforward assume that Υ_0 is a set that is a union of: a subset of X^h and a subfamily of edges (treated as closed segments) of the triangulation. Then, with $\mathrm{X}^h_0 := \mathrm{X}^h \cap \Upsilon_0$, the space $\mathcal{U}^h(\Omega)$ is isomorphic to the $2m$-dimensional space of nodal displacements at vertices $\mathrm{X}^h \setminus \mathrm{X}^h_0$, where $m = \operatorname{card}(\mathrm{X}^h \setminus \mathrm{X}^h_0)$. We introduce an ordering of the vertices: $\{v_1, \ldots, v_m\} = \mathrm{X}^h \setminus \mathrm{X}^h_0$ and a unique interpolation operator $\iota^h : \mathbb{R}^{m\times 2} \to \mathcal{U}^h(\Omega)$ being the isomorphism that for $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^m$ satisfies

$$u = \iota^h(\mathbf{u}_1, \mathbf{u}_2) \qquad \Leftrightarrow \qquad u(v_j) = \mathbf{u}_1(j)\,\mathrm{e}_1 + \mathbf{u}_2(j)\,\mathrm{e}_2 \qquad \forall\, j \in \{1, \ldots, m\}.$$

Consequently, in the context of problem $\mathcal{P}_h$ the load F may be reduced to vectors $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{R}^m$ that are uniquely defined via:

$$\langle \mathbf{u}_1, \mathbf{F}_1\rangle + \langle \mathbf{u}_2, \mathbf{F}_2\rangle = \int_{\overline{\Omega}} \big\langle \iota^h(\mathbf{u}_1, \mathbf{u}_2), F\big\rangle \qquad \forall\, \mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^m. \tag{4.45}$$

For any function $u \in \mathcal{U}^h(\Omega) \subset \mathrm{Lip}(\Omega; \mathbb{R}^2)$ the strain $e(u)$ exists $\mathcal{L}^2$-a.e. in the classic sense due to Rademacher theorem, cf. [82]. Clearly $e(u) : \Omega \to \mathcal{S}^{2\times 2}$ is constant a.e. in a fixed element E^h_i, which will be fundamental for the numerical method herein proposed. We define an isomorphism $\chi : \mathbb{R}^3 \to \mathcal{S}^{2\times 2}$ that is given via the formula below:

$$\chi\big(a^{11}, a^{22}, a^{12}\big) := a^{11}\,\mathrm{e}_1 \otimes \mathrm{e}_1 + a^{22}\,\mathrm{e}_2 \otimes \mathrm{e}_2 + \sqrt{2}\,a^{12}\,\mathrm{e}_1 \overset{s}{\otimes} \mathrm{e}_2.$$

The map χ is an isometry, namely

$$\Big\langle \big(a^{11}, a^{22}, a^{12}\big), \big(b^{11}, b^{22}, b^{12}\big)\Big\rangle_{\mathbb{R}^3} = \Big\langle \chi\big(a^{11}, a^{22}, a^{12}\big), \chi\big(b^{11}, b^{22}, b^{12}\big)\Big\rangle_{\mathcal{S}^{2\times 2}}. \tag{4.46}$$

Next, we build six *geometric matrices* $\mathbf{B}^{k_1k_2}_{k_3} \in \mathbb{R}^{n\times m}$, where $k_1, k_2, k_3 \in \{1, 2\}$ and $k_1 \le k_2$, that are uniquely defined through the following relation: for all vectors $\boldsymbol{\epsilon}^{11}, \boldsymbol{\epsilon}^{22}, \boldsymbol{\epsilon}^{12} \in \mathbb{R}^n$ there holds for every element index $i \in \{1, \ldots, n\}$

$$e\big(\iota^h(\mathbf{u}_1, \mathbf{u}_2)\big) = \chi\big(\boldsymbol{\epsilon}^{11}(i), \boldsymbol{\epsilon}^{22}(i), \boldsymbol{\epsilon}^{12}(i)\big) \quad \text{a.e. in } E^h_i \qquad \Leftrightarrow \qquad \begin{cases} \boldsymbol{\epsilon}^{11} = \mathbf{B}^{11}_1\mathbf{u}_1 + \mathbf{B}^{11}_2\mathbf{u}_2 \\ \boldsymbol{\epsilon}^{22} = \mathbf{B}^{22}_1\mathbf{u}_1 + \mathbf{B}^{22}_2\mathbf{u}_2 \\ \boldsymbol{\epsilon}^{12} = \mathbf{B}^{12}_1\mathbf{u}_1 + \mathbf{B}^{12}_2\mathbf{u}_2 \end{cases} \tag{4.47}$$

The key idea behind this finite element method consists in observing that the set

$$\mathcal{K}^4_\rho := \Big\{(r, \epsilon^{11}, \epsilon^{22}, \epsilon^{12}) \in \mathbb{R}^4 \,:\, \rho\Big(\chi\big(\epsilon^{11}, \epsilon^{22}, \epsilon^{12}\big)\Big) \le r\Big\}$$

is a closed convex cone, which paves the way for rewriting $(\mathcal{P}_h)$ as a conic programming problem, cf. Section 2.6. Beforehand we characterize the dual cone $\big(\mathcal{K}^4_\rho\big)^*$ (cf. (2.30)) by means of polar function ρ^0:

Proposition 4.46. *The cone dual to $\mathcal{K}^4_\rho$ enjoys the following characterization:*

$$\big(r^0,-\sigma^{11},-\sigma^{22},-\sigma^{12}\big)\in\big(\mathcal{K}^4_\rho\big)^* \qquad\Leftrightarrow\qquad \big(r^0,\sigma^{11},\sigma^{22},\sigma^{12}\big)\in\mathcal{K}^4_{\rho^0}$$

where

$$\mathcal{K}^4_{\rho^0}=\Big\{(r^0,\sigma^{11},\sigma^{22},\sigma^{12})\in\mathbb{R}^4\,:\,\rho^0\Big(\chi\big(\sigma^{11},\sigma^{22},\sigma^{12}\big)\Big)\le r^0\Big\}.$$

Proof. Since χ is a linear isometry, cf. (4.46) we observe that

$$\big\langle\,(r,\epsilon^{11},\epsilon^{22},\epsilon^{12})\,,\,(r^0,-\sigma^{11},-\sigma^{22},-\sigma^{12})\,\big\rangle_{\mathbb{R}^4}=r\,r^0-\langle\xi,\sigma\rangle_{\mathcal{S}^{2\times2}}$$

where $\xi=\chi\big(\epsilon^{11},\epsilon^{22},\epsilon^{12}\big)$ and $\sigma=\chi\big(\sigma^{11},\sigma^{22},\sigma^{12}\big)$. The problem thus reduces to determining necessary and sufficient condition for r^0,σ for which $\langle\xi,\sigma\rangle\le r\,r^0$ for any r,ξ satisfying $\rho(\xi)\le r$. This conditions precisely reads: $\rho^0(\sigma)\le r^0$ which furnishes the claim. □

By gathering all of the foregoing results the problem $(\mathcal{P}_h)$ can be readily reformulated as a conic programming problem:

$$\sup_{\substack{\mathbf{u}_1,\,\mathbf{u}_2\in\mathbb{R}^m\\ \boldsymbol{\epsilon}^{11},\,\boldsymbol{\epsilon}^{22},\,\boldsymbol{\epsilon}^{12}\in\mathbb{R}^n,\\ \mathbf{r}\in\mathbb{R}^n}}\left\{\langle\mathbf{u}_1,\mathbf{F}_1\rangle+\langle\mathbf{u}_2,\mathbf{F}_2\rangle\,:\,\begin{array}{c}\mathbf{B}_1^{11}\mathbf{u}_1+\mathbf{B}_2^{11}\mathbf{u}_2-\boldsymbol{\epsilon}^{11}=\mathbf{0}\\ \mathbf{B}_1^{22}\mathbf{u}_1+\mathbf{B}_2^{22}\mathbf{u}_2-\boldsymbol{\epsilon}^{22}=\mathbf{0}\\ \mathbf{B}_1^{12}\mathbf{u}_1+\mathbf{B}_2^{12}\mathbf{u}_2-\boldsymbol{\epsilon}^{12}=\mathbf{0}\\ \mathbf{r}=\mathbf{1}\\ \big(\mathbf{r}(i),\boldsymbol{\epsilon}^{11}(i),\boldsymbol{\epsilon}^{22}(i),\boldsymbol{\epsilon}^{12}(i)\big)\in\mathcal{K}^4_\rho\qquad\forall\,i\end{array}\right\}\qquad(\mathbf{P}_h)$$

where $\mathbf{1}\in\mathbb{R}^n$ stands for the vector whose all components are equal to 1. Below we shall show that the dual of $(\mathbf{P}_h)$ reads

$$\inf_{\substack{\boldsymbol{\sigma}^{11},\,\boldsymbol{\sigma}^{22},\,\boldsymbol{\sigma}^{12}\in\mathbb{R}^n,\\ \mathbf{r}^0\in\mathbb{R}^n}}\left\{\big\langle\mathbf{1},\mathbf{r}^0\big\rangle\,:\,\begin{array}{c}(\mathbf{B}_1^{11})^\top\boldsymbol{\sigma}^{11}+(\mathbf{B}_1^{22})^\top\boldsymbol{\sigma}^{22}+(\mathbf{B}_1^{12})^\top\boldsymbol{\sigma}^{12}=\mathbf{F}_1\\ (\mathbf{B}_2^{11})^\top\boldsymbol{\sigma}^{11}+(\mathbf{B}_2^{22})^\top\boldsymbol{\sigma}^{22}+(\mathbf{B}_2^{12})^\top\boldsymbol{\sigma}^{12}=\mathbf{F}_2\\ \big(\mathbf{r}^0(i),\boldsymbol{\sigma}^{11}(i),\boldsymbol{\sigma}^{22}(i),\boldsymbol{\sigma}^{12}(i)\big)\in\mathcal{K}^4_{\rho^0}\qquad\forall\,i\end{array}\right\}\qquad(\mathbf{P}_h^*)$$

Theorem 4.47. *Let us assume that $\int\langle u_0,F\rangle=0$ for any $u_0\in\mathcal{R}$ and, moreover,* $(\mathrm{A}_{\mathrm{ellip}})$ *holds. Then*

$$Z_h=\max\mathbf{P}_h=\min\mathbf{P}_h^*<\infty,$$

in particular there exist solutions of both problems.

Moreover, the variables $(\mathbf{u}_1,\mathbf{u}_2,\boldsymbol{\epsilon}^{11},\boldsymbol{\epsilon}^{22},\boldsymbol{\epsilon}^{12},\mathbf{r})$ and $(\boldsymbol{\sigma}^{11},\boldsymbol{\sigma}^{22},\boldsymbol{\sigma}^{12},\mathbf{r}^0)$ that are admissible for problems $(\mathbf{P}_h)$ and $(\mathbf{P}_h^*)$, respectively, are solutions of those problems if and only if for any $i\in\{1,\dots,n\}$ there hold optimality conditions

(i) $\langle\epsilon_i,\sigma_i\rangle=\rho^0(\sigma_i)$,
(ii) $\mathbf{r}^0(i)=\rho^0(\sigma_i)$,

where $\epsilon_i=\chi\big(\boldsymbol{\epsilon}^{11}(i),\boldsymbol{\epsilon}^{22}(i),\boldsymbol{\epsilon}^{12}(i)\big)\in\mathcal{S}^{2\times2}$ and $\sigma_i=\chi\big(\boldsymbol{\sigma}^{11}(i),\boldsymbol{\sigma}^{22}(i),\boldsymbol{\sigma}^{12}(i)\big)\in\mathcal{S}^{2\times2}$.

Proof. We prove the result by putting the pair of problems $(\mathbf{P}_h)$ and $(\mathbf{P}_h^*)$ in the framework of the canonical conic programming problem presented in Section 2.6: the problem $(\mathbf{P}_h^*)$ becomes the primal problem (CP) given therein, while $(\mathbf{P}_h)$ the dual problem (CP*) (up to a small modification, see below) provided that we set:

$$\mathbf{x} = \begin{bmatrix} \mathbf{r}^0 \\ \boldsymbol{\sigma}^{11} \\ \boldsymbol{\sigma}^{22} \\ \boldsymbol{\sigma}^{12} \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B}_1^{11} & \mathbf{B}_2^{11} \\ \mathbf{B}_1^{22} & \mathbf{B}_2^{22} \\ \mathbf{B}_1^{12} & \mathbf{B}_2^{12} \end{bmatrix}^\top, \quad \mathbf{b} = \begin{bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \boldsymbol{\eta}_i = \begin{bmatrix} \mathbf{r}(i) \\ \tilde{\boldsymbol{\epsilon}}^{11}(i) \\ \tilde{\boldsymbol{\epsilon}}^{22}(i) \\ \tilde{\boldsymbol{\epsilon}}^{12}(i) \end{bmatrix};$$

$\mathbf{Q}_i \in \mathbb{R}^{4\times 4n}$ for $i \in \{1,\dots,n\}$ are 0 and 1 matrices with non-zero entries as follows: $\mathbf{Q}_i\big(1,i\big) = \mathbf{Q}_i(2,2i) = \mathbf{Q}_i(3,3i) = \mathbf{Q}_i(4,4i) = 1$; for each i we choose the cone: $\mathcal{K}_i = \mathcal{K}^4_{\rho^0}$. The problem $(\mathbf{P}_h^*)$ is thus recast by (CP) directly, whilst (CP*) requires alteration of signs: by Proposition 4.46 (applied to ρ in place of ρ^0) the cone constraint $\big(\mathbf{r}(i), \tilde{\boldsymbol{\epsilon}}^{11}(i), \tilde{\boldsymbol{\epsilon}}^{22}(i), \tilde{\boldsymbol{\epsilon}}^{12}(i)\big) \in \mathcal{K}_i^*$ can be rewritten as $\big(\mathbf{r}(i), -\tilde{\boldsymbol{\epsilon}}^{11}(i), -\tilde{\boldsymbol{\epsilon}}^{22}(i), -\tilde{\boldsymbol{\epsilon}}^{12}(i)\big) \in \mathcal{K}^4_\rho$. Therefore, the problem $(\mathbf{P}_h)$ is recovered by substitution: $\boldsymbol{\epsilon}^{11} = -\tilde{\boldsymbol{\epsilon}}^{11}$, $\boldsymbol{\epsilon}^{22} = -\tilde{\boldsymbol{\epsilon}}^{22}$, $\boldsymbol{\epsilon}^{12} = -\tilde{\boldsymbol{\epsilon}}^{12}$.

Next, we verify assumptions (1), (2) in Definition 2.70. It is straightforward to check that $\bigcap_{i=1}^n \operatorname{Ker} \mathbf{Q}_i = \{0\} \subset \mathbb{R}^{4n}$. By definition of the geometric matrices $\mathbf{B}_{k_3}^{k_1k_2}$ we find that the prerequisite for the load F is equivalent to $\mathbf{b} \perp \operatorname{Ker} \mathbf{A}^\top$, i.e. $\mathbf{b} \in \operatorname{Im} \mathbf{A}$. Therefore all the linearly dependent rows of $\mathbf{A}$ may be deleted together with the corresponding entries in $\mathbf{b}$ and $\mathbf{y}$. Assumptions (1), (2) are readily verified for the modified problem, therefore in the sequel of the proof it is not restrictive to assume that rows of $\mathbf{A}$ are linearly independent.

To show the first part of the assertion we shall point to strictly feasible variables $\mathbf{x}$ and $(\mathbf{y}, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n)$ in the sense of Definition 2.70. For (CP) let $\mathbf{x} = (\mathbf{r}^0, \boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{22}, \boldsymbol{\sigma}^{12})$ be any solution of $\mathbf{A}\mathbf{x} = \mathbf{b}$ and we observe that $\mathbf{r}^0$ may be modified freely – hence we choose $\mathbf{r}^0(i) = \varepsilon + \rho^0\big(\chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)\big)$ where $\varepsilon > 0$. Owing to assumption $(\mathrm{A}_{\mathrm{ellip}})$ we are certain that $\mathbf{r}^0(i) < \infty$ (cf. Proposition 3.13) therefore $\big(\mathbf{r}^0(i), \boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big) \in \operatorname{int} \mathcal{K}^4_{\rho^0}$ which establishes strict feasibility of $\mathbf{x}$ and guarantees that $0 \le \inf \mathrm{CP} < \infty$. By virtue of statements (i), (ii) in Theorem 2.71 equality $\sup \mathrm{CP}^* = \inf \mathrm{CP}$ holds true and (CP*) attains a solution. A strictly feasible solution for (CP*) is trivial to find: we set $\mathbf{y} = \mathbf{0}$ and $\boldsymbol{\eta}_i = (1,0,0,0)$ for each i. According to (iii) in Theorem 2.71 existence of solution for (CP) follows as well.

To prove the "moreover part" of the result we employ statement (iv) in Theorem 2.71: the vectors $\mathbf{x}$ and $(\mathbf{y}, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n)$ that are feasible for (CP) and (CP*), respectively, solve those problem if and only if $\langle\, \mathbf{Q}_i\mathbf{x}\,,\, \boldsymbol{\eta}_i \,\rangle = 0$, i.e.

$$\begin{aligned} 0 &= \Big\langle \big(\mathbf{r}^0(i), \boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)\,,\, \big(\mathbf{r}(i), \tilde{\boldsymbol{\epsilon}}^{11}(i), \tilde{\boldsymbol{\epsilon}}^{22}(i), \tilde{\boldsymbol{\epsilon}}^{12}(i)\big) \Big\rangle_{\mathbb{R}^4} \\ &= \Big\langle \big(\mathbf{r}^0(i), \boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)\,,\, \big(1, -\boldsymbol{\epsilon}^{11}(i), -\boldsymbol{\epsilon}^{22}(i), -\boldsymbol{\epsilon}^{12}(i)\big) \Big\rangle_{\mathbb{R}^4} \\ &= \mathbf{r}^0(i) - \langle \epsilon_i, \sigma_i \rangle_{\mathcal{S}^{2\times 2}} \end{aligned} \tag{4.48}$$

where once again we acknowledged that χ is an isometry, see (4.46). From the conic constraints we find that $\rho(\epsilon_i) \le 1$ and $\rho^0(\sigma_i) \le \mathbf{r}^0(i)$ and the asserted optimality conditions (i), (ii) follow by (4.48). □

With the equivalence between problems $(\mathcal{P}_h)$ and $(\mathbf{P}_h)$ it is natural to attempt tracing the dual conic program $(\mathbf{P}_h^*)$ back to the problem $(\mathcal{P}_h^*)$. We start by assigning vectors $\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{22}, \boldsymbol{\sigma}^{12} \in \mathbb{R}^n$ to any stress function $\sigma \in L^1(\Omega; \mathcal{S}^{2\times 2})$ as below:

$$\begin{cases} \boldsymbol{\sigma}^{11}(i) = \int_{E_i^h} \langle \sigma, \mathrm{e}_1 \otimes \mathrm{e}_1 \rangle \, d\mathcal{L}^2, \\ \boldsymbol{\sigma}^{22}(i) = \int_{E_i^h} \langle \sigma, \mathrm{e}_2 \otimes \mathrm{e}_2 \rangle \, d\mathcal{L}^2, \\ \boldsymbol{\sigma}^{12}(i) = \int_{E_i^h} \sqrt{2} \, \langle \sigma, \mathrm{e}_1 \overset{s}{\otimes} \mathrm{e}_2 \rangle \, d\mathcal{L}^2. \end{cases} \tag{4.49}$$

We thus arrive at an algebraic representation of the set $\Sigma^h(\Omega, F)$:

Proposition 4.48. *A function $\sigma \in L^1(\Omega; \mathcal{S}^{2\times 2})$ is an element of $\Sigma^h(\Omega, F)$ if and only if*

$$\begin{cases} (\mathbf{B}_1^{11})^\top \boldsymbol{\sigma}^{11} + (\mathbf{B}_1^{22})^\top \boldsymbol{\sigma}^{22} + (\mathbf{B}_1^{12})^\top \boldsymbol{\sigma}^{12} = \mathbf{F}_1 \\ (\mathbf{B}_2^{11})^\top \boldsymbol{\sigma}^{11} + (\mathbf{B}_2^{22})^\top \boldsymbol{\sigma}^{22} + (\mathbf{B}_2^{12})^\top \boldsymbol{\sigma}^{12} = \mathbf{F}_2 \end{cases} \tag{4.50}$$

where $\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{22}, \boldsymbol{\sigma}^{12} \in \mathbb{R}^n$ are given by (4.49).

Proof. By using the isomorphism ι^h function $\sigma \in L^1(\Omega; \mathcal{S}^{d\times d})$ belongs to $\Sigma^h(\Omega, F)$ if and only if

$$\int_\Omega \big\langle e\big(\iota^h(\mathbf{u}_1, \mathbf{u}_2)\big), \sigma \big\rangle \, d\mathcal{L}^2 = \int_{\overline{\Omega}} \big\langle \iota^h(\mathbf{u}_1, \mathbf{u}_2), F \big\rangle \qquad \forall\, \mathbf{u}_1, \mathbf{u_2} \in \mathbb{R}^m$$

which is equivalent to

$$\sum_{i=1}^n \int_{E_i^h} \big\langle \chi\big(\boldsymbol{\epsilon}^{11}(i), \boldsymbol{\epsilon}^{22}(i), \boldsymbol{\epsilon}^{12}(i)\big) \, , \, \sigma \big\rangle \, d\mathcal{L}^2 = \langle \mathbf{u}_1, \mathbf{F}_1 \rangle + \langle \mathbf{u}_2, \mathbf{F}_2 \rangle \qquad \forall\, \mathbf{u}_1, \mathbf{u_2} \in \mathbb{R}^m \tag{4.51}$$

where $\boldsymbol{\epsilon}^{11}$, $\boldsymbol{\epsilon}^{22}$, $\boldsymbol{\epsilon}^{12}$ are expressed by $\mathbf{u}_1, \mathbf{u_2}$ according to (4.47). For each i the integral over E_i^h above reduces to $\big\langle \chi\big(\boldsymbol{\epsilon}^{11}(i), \boldsymbol{\epsilon}^{22}(i), \boldsymbol{\epsilon}^{12}(i)\big) \, , \, \chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big) \big\rangle$ therefore, by (4.46), equation (4.51) can be rewritten as

$$\big\langle \boldsymbol{\epsilon}^{11}, \boldsymbol{\sigma}^{11} \big\rangle + \big\langle \boldsymbol{\epsilon}^{22}, \boldsymbol{\sigma}^{22} \big\rangle + \big\langle \boldsymbol{\epsilon}^{12}, \boldsymbol{\sigma}^{12} \big\rangle = \langle \mathbf{u}_1, \mathbf{F}_1 \rangle + \langle \mathbf{u}_2, \mathbf{F}_2 \rangle \qquad \forall\, \mathbf{u}_1, \mathbf{u_2} \in \mathbb{R}^m$$

or

$$\begin{aligned} &\big\langle \mathbf{B}_1^{11}\mathbf{u}_1, \boldsymbol{\sigma}^{11} \big\rangle + \big\langle \mathbf{B}_1^{22}\mathbf{u}_1, \boldsymbol{\sigma}^{22} \big\rangle + \big\langle \mathbf{B}_1^{12}\mathbf{u}_1, \boldsymbol{\sigma}^{12} \big\rangle \\ &+ \big\langle \mathbf{B}_2^{11}\mathbf{u}_2, \boldsymbol{\sigma}^{11} \big\rangle + \big\langle \mathbf{B}_2^{22}\mathbf{u}_2, \boldsymbol{\sigma}^{22} \big\rangle + \big\langle \mathbf{B}_2^{12}\mathbf{u}_2, \boldsymbol{\sigma}^{12} \big\rangle = \langle \mathbf{u}_1, \mathbf{F}_1 \rangle + \langle \mathbf{u}_2, \mathbf{F}_2 \rangle \qquad \forall\, \mathbf{u}_1, \mathbf{u_2} \in \mathbb{R}^m \end{aligned}$$

and the passage to algebraic equations (4.50) may be considered standard. □

The result below shows how to reconstruct solutions of $(\mathcal{P}_h)$ and $(\mathcal{P}_h^*)$ based on solutions of the conic programs.

Corollary 4.49. *Let $(\mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\epsilon}^{11}, \boldsymbol{\epsilon}^{22}, \boldsymbol{\epsilon}^{12}, \mathbf{r})$ and $(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{22}, \boldsymbol{\sigma}^{12}, \mathbf{r}^0)$ be solutions of problems $(\mathbf{P}_h)$ and $(\mathbf{P}_h^*)$, respectively. We choose the unique functions $u_h \in \mathcal{U}^h(\Omega)$ and $\sigma_h \in L^1(\Omega; \mathcal{S}^{2\times 2})$ satisfying*

$$u_h = \iota^h(\mathbf{u}_1, \mathbf{u}_2), \qquad \sigma_h = \frac{1}{|E_i^h|} \, \chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big) \quad \text{a.e. in } E_i^h.$$

Then, functions u_h and σ_h solve problems $(\mathcal{P}_h)$ and $(\mathcal{P}_h^*)$, respectively, while $\min \mathcal{P}_h^* = Z_h$.

Proof. The problem $(\mathbf{P}_h)$ is equivalent to the already finite dimensional problem $(\mathcal{P}_h)$ therefore optimality of u_h is clear. It is easy to check that the element-wise constant function σ_h and vectors $\boldsymbol{\sigma}^{11}$, $\boldsymbol{\sigma}^{22}$, $\boldsymbol{\sigma}^{12}$ satisfy relation (4.49) thus $\sigma_h \in \Sigma^h(\Omega, F)$. By utilizing optimality condition (ii) in Theorem 4.47 we obtain

$$\int_\Omega \rho^0(\sigma_h)\, d\mathcal{L}^2 = \sum_{i=1}^n \frac{1}{|E_i^h|} \int_{E_i^h} \rho^0\Big(\chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)\Big)\, d\mathcal{L}^2 = \sum_{i=1}^n \mathbf{r}^0(i) = \big\langle \mathbf{1}, \mathbf{r}^0 \big\rangle = Z_h$$

rendering σ_h optimal for $(\mathcal{P}_h^*)$ since $Z_h = \max \mathcal{P}_h \leq \inf \mathcal{P}_h^*$. □

The finite element method that is put forward approximates the pair of problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$. It is the author's hope that with a *regular* (cf. [59, Chapter 3, Section 3.1]) family of triangulations $\mathcal{T}^h$ solutions u_h and σ_h computed in accordance with Corollary 4.49 converge to solutions $u \in \overline{\mathcal{U}}_1$ and $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{2\times 2})$ of problems $(\overline{\mathcal{P}})$ and $(\mathcal{P}^*)$ when h approaches zero, more accurately: $u_h \to u$ uniformly in Ω, $e(u_h) \overset{*}{\rightharpoonup} e(u)$ in $L^\infty(\Omega; \mathcal{S}^{2\times 2})$ and $\sigma_h\, \mathcal{L}^2 \llcorner \Omega \overset{*}{\rightharpoonup} \varsigma$. As already mentioned, this analysis lies beyond the scope of this dissertation.

We shall address a more practical question: how from functions u_h, σ_h does one recover an approximate optimal constitutive field $\lambda = \lambda_h$? Since the relative stress σ_h induces the absolute stress $\varsigma_h = \sigma_h\, \mathcal{L}^2 \llcorner \Omega$ we may construct λ_h by using an adapted version of Corollary 4.7 (in fact, we only need to replace Z with Z_h). Below we will validate this idea by showing that λ_h thus constructed indeed solves a suitably defined approximate Free Material Design problem (FMD_h). We observe that from the FE method we always obtain measures that are absolutely continuous with respect to Lebesgue measure – it is thus natural to represent the constitutive fields as $\lambda = \mathscr{H}\, \mathcal{L}^2 \llcorner \Omega$ or, ultimately, to speak of the function $\mathscr{H}$ only.

We start by defining the approximate elastic compliance which for a given triangulation $\mathcal{T}^h$ and any anisotropy function $\mathscr{H} \in L^1(\Omega; \mathcal{K})$ reads

$$\mathcal{C}_h(\mathscr{H}) := \sup \left\{ \int_{\overline{\Omega}} \langle u, F \rangle - \int_\Omega j\big(\mathscr{H}, e(u)\big)\, d\mathcal{L}^2 \ : \ u \in \mathcal{U}^h(\Omega) \right\}. \tag{4.52}$$

The maximization problem above can be recognized as the classical FEM approximation of elasticity problem, cf. [59]. By the approximate Free Material Design problem we shall understand the problem with approximate objective functional, instead of approximating the set of admissible constitutive fields:

$$\mathcal{C}_{h,\min} = \inf \left\{ \mathcal{C}_h(\mathscr{H}) \ : \ \mathscr{H} \in L^1(\Omega; \mathcal{K}),\ \int_\Omega c(\mathscr{H})\, d\mathcal{L}^2 \leq C_0 \right\} \tag{FMD_h}$$

Once we assume that $\mathscr{H} \in L^\infty(\Omega; \mathcal{K})$, i.e. $\mathscr{H}$ is essentially bounded, an almost direct use of the arguments from the proof of Proposition 4.4 for $\mu = \mathcal{L}^2 \llcorner \Omega$ yields

$$\mathcal{C}_h(\mathscr{H}) = \inf \left\{ \int_\Omega j^*\big(\mathscr{H}, \sigma\big)\, d\mathcal{L}^2 \ : \ \sigma \in \Sigma^h(\Omega, F) \right\}. \tag{4.53}$$

With all the results that were established in this section we could easily adapt some of the arguments used in Sections 4.2 and 4.3 that would furnish counterparts of Lemma 4.2, Theorem 4.6, Proposition 4.13 in the approximate FEM setting. Since there would be no new steps involved we shall not display all the modified proofs and instead we give the final result straightaway:

Theorem 4.50. *Let* $(\mathbf{u}_1, \mathbf{u}_2, \boldsymbol{\epsilon}^{11}, \boldsymbol{\epsilon}^{22}, \boldsymbol{\epsilon}^{12}, \mathbf{r})$ *and* $(\boldsymbol{\sigma}^{11}, \boldsymbol{\sigma}^{22}, \boldsymbol{\sigma}^{12}, \mathbf{r}^0)$ *be solutions of problems* $(\mathbf{P}_h)$ *and* $(\mathbf{P}_h^*)$*, respectively. For each* i *let* $H_i \in \mathcal{K}_1$ *be any Hooke tensor such that*

$$H_i \in \underline{\mathcal{K}}_1\Big(\chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)\Big).$$

Then, the Hooke field $\mathscr{H}_h \in L^\infty(\Omega; \mathcal{K})$ *satisfying*

$$\mathscr{H}_h = \frac{1}{|E_i^h|}\,\frac{C_0}{Z_h}\,\mathbf{r}^0(i)\,H_i \qquad \text{a.e. in } E_i^h \tag{4.54}$$

solves the approximate optimal design problem (FMD_h)*. Moreover, the displacement function* $u_h \in \mathcal{U}^h(\Omega)$ *and stress function* $\sigma_h \in L^\infty(\Omega; \mathcal{S}^{2\times 2})$ *satisfying*

$$u_h = \left(\frac{Z_h}{C_0}\right)^{p'-1} \iota^h(\mathbf{u}_1, \mathbf{u}_2), \qquad \sigma_h = \frac{1}{|E_i^h|}\,\chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big) \quad \text{a.e. in } E_i^h$$

solve the approximate elasticity problems (4.52) *and, respectively,* (4.53) *for* $\mathscr{H} = \mathscr{H}_h$*.*

Before moving on to numerical implementation of the developed finite element method we give an idea for analysis of convergence of approximate solutions $\mathscr{H}_h$ given by (4.54). We stress that $\mathscr{H}_h$ thus constructed is an exact solution of problem (FMD_h) where only the objective functional $\mathcal{C}_h$ depends on h. Study of convergence of the sequence $\{\mathscr{H}_h\}$ therefore reduces to study of Γ-convergence of $\mathcal{C}_h$, cf. [47]. To be more rigorous we must go back to functionals on the space of measures, i.e. for any $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ we define

$$\overline{\mathcal{C}}_h(\lambda) = \begin{cases} \mathcal{C}_h\left(\frac{d\lambda}{d\mathcal{L}^2}\right) & \text{if } \lambda \ll \mathcal{L}^2 \llcorner \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

With all the assumptions made in this section in force, we put forward the following conjecture:

Conjecture 4.51. *If* $\mathcal{T}^h$ *is a regular family of triangulations the sequence of functionals* $\overline{\mathcal{C}}_h : \mathcal{M}(\overline{\Omega}; \mathcal{K}) \to \overline{\mathbb{R}}$ Γ*-converges to* $\mathcal{C} : \mathcal{M}(\overline{\Omega}; \mathcal{K}) \to \overline{\mathbb{R}}$ *given by* (4.20) *when* $h \to 0$*.*

As a result, with $\mathscr{H}_h$ *computed in accordance with Theorem 4.50*

$$\mathscr{H}_h\,\mathcal{L}^2 \llcorner \Omega \overset{*}{\rightharpoonup} \lambda \qquad \text{in } \mathcal{M}(\overline{\Omega}; \mathcal{K})$$

while λ *solves the* (FMD) *problem for two dimensional elasticity.*

4.6.2. Numerical simulations

4.6.2.1. The Anisotropic Material Design setting – implementation as a conic quadratic programming problem

The finite element method for the Free Material Design of 2D elastic bodies was implemented in MATLAB® R2018b. The simplest case of rectangular feasible domain was accounted for only, allowing a regular finite element mesh. The set of vertices X^h of the triangulation $\mathcal{T}^h$ is a grid of points regularly spaced in two orthogonal directions parallel to the edges of Ω; each rectangle connecting the adjacent vertices is then divided by one of its diagonals (whose direction changes depending on the quadrant of Ω to preserve symmetry) to arrive at a mesh of triangular elements.

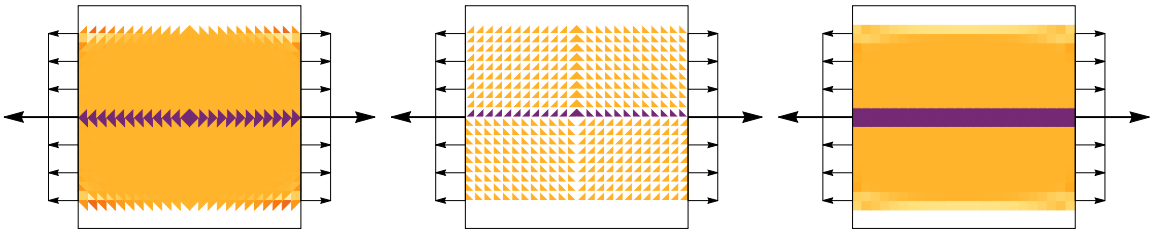
The pair of mutually dual conic programming problems $(\mathbf{P}_h)$, $(\mathbf{P}_h^*)$ itself is tackled by using a dedicated solver from the MOSEK® 9.2 toolbox, cf. [10]. The solver employs a highly efficient interior-point algorithm that is suited for large scale problems, cf. [9]. The MOSEK toolbox offers several standard types of conic constraints, namely those induced by *quadratic*, *power* or *exponential* cones. This clearly limits the use of the software for general triples $j, \mathcal{K}, c$ that induce general gauges ρ, ρ^0. In fact, from the anisotropy design settings listed in Section 3.3, only for the AMD setting where $\rho = \rho^0 = |\cdot|$ the cones $\mathcal{K}^4_\rho, \mathcal{K}^4_{\rho^0}$ can be directly written as the quadratic cones (see (2.31) in Section 2.6), i.e.

$$\mathcal{K}^4_\rho = \mathcal{K}^4_{\rho^0} = \mathcal{K}^4_{\text{quad}} = \Big\{(r^0, \sigma^{11}, \sigma^{22}, \sigma^{12}) \in \mathbb{R}^4 \,:\, r^0 \geq \sqrt{(\sigma^{11})^2 + (\sigma^{22})^2 + (\sigma^{12})^2}\Big\}.$$

Notwithstanding this, for each anisotropy design setting herein considered we will be able to rewrite $(\mathbf{P}_h)$, $(\mathbf{P}_h^*)$ as a pair of *conic quadratic programs*, i.e. conic programs with conic quadratic constraints only. For the time being we stay within the AMD framework where based on solution of problem $(\mathbf{P}_h^*)$ the optimal distribution $\mathscr{H}_h \in L^1\big(\Omega; \mathscr{L}_+(\mathcal{S}^{2\times 2})\big)$ of Hooke tensor reads

$$\mathscr{H}_h = \frac{1}{|E_i^h|}\frac{C_0}{Z_h}\,\mathbf{r}^0(i)\,\frac{\chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)}{\mathbf{r}^0(i)} \otimes \frac{\chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)}{\mathbf{r}^0(i)} \qquad \text{a.e. in } E_i^h,$$

where we used the fact that $\mathbf{r}^0(i) = \rho^0\big(\chi\big(\boldsymbol{\sigma}^{11}(i), \boldsymbol{\sigma}^{22}(i), \boldsymbol{\sigma}^{12}(i)\big)\big)$ and characterization (3.19) of the set $\underline{\mathcal{K}}_1(\sigma)$. We can see that $\operatorname{Tr}\mathscr{H}_h = \frac{1}{|E_i^h|}\frac{C_0}{Z_h}\,\mathbf{r}^0(i)$.

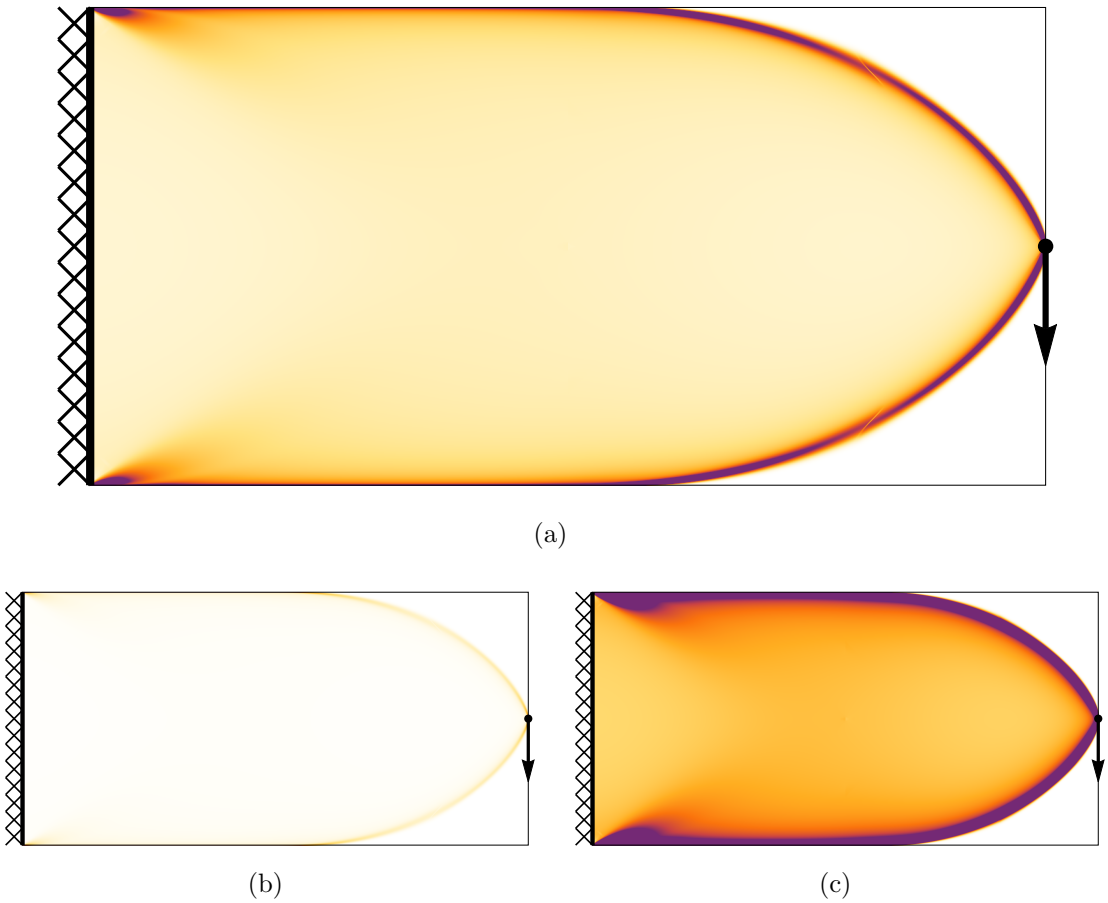


(a) (b) (c)

Figure 4.2: Numerical simulations of the optimal plate design via FEM, map of $\operatorname{Tr}\mathscr{H}_h$: (a) original FE solution; (b) FE solution of a perturbed problem; (c) averaged material distribution.

Prior to solving the finite element problems for fine meshes and different anisotropy settings we perform numerical simulation for the uni-axially stressed plate problem (solved analytically in Example 4.34) for a coarse FE mesh. We consider the AMD setting for which the pair $(\mathbf{P}_h)$, $(\mathbf{P}_h^*)$ is originally a conic quadratic program. For a quadratic design domain $\Omega = (-1/2, 1/2)^2$ we take a mesh of 25×25 nodes furnishing 1152 triangular elements; we fix $Q = q = 1$ and $b = 3/4$, cf. Fig. 4.1. In Fig. 4.2(a) we present the solution in the form of a heat-map of $\operatorname{Tr}\mathscr{H}_h$; the darker the colour the higher the value of $\operatorname{Tr}\mathscr{H}_h$. One may observe that the middle horizontal bar is portrayed by a saw-wave-like shape and similar shapes occur along the edges of the elastic body. What is more, in vicinity of corners elements of high and low material density intertwine. These numerical effects may be recognized as a *checker-board phenomenon* which is typical for optimal design when

using FE methods with low-order elements (herein displacement function u is element-wise affine). In the literature on optimal design it is more common to use mesh of quads where due to this effect the optimal material distribution resembles a chess board – hence the name of the effect; cf. [4, Section 5.2.3] for more details on the issue. We add that $Z_h = 1.75$ which is precisely equal to the analytical value $Z = Q\,a + q\,ab$ which means that u_h is the analytical solution of $(\overline{\mathcal{P}})$.



(a)

(b) (c)

Figure 4.3: The cantilever problem in the Anisotropic Material Design setting – distribution of $\mathrm{Tr}\mathscr{H}_h$ optimal in (FMD_h): heat-map of $\min\{\mathrm{Tr}\mathscr{H}_h, \alpha\,\|\mathrm{Tr}\mathscr{H}_h\|_{L^\infty}\}$ for (a) $\alpha = 0.005$; (b) for $\alpha = 0.05$; (c) for $\alpha = 0.0015$.

In order to amplify the checker-board effect we will solve a perturbed FE problem. Instead of constant vector $\mathbf{1}$ in $(\mathbf{P}_h)$, $(\mathbf{P}_h^*)$ we take vector $\tilde{\mathbf{1}}$ defined as follow: $\tilde{\mathbf{1}}(i) = 1$ for i corresponding to the lower triangular elements (lower with respect to the diagonal that splits the quads) and $\tilde{\mathbf{1}}(i) = 1 + 10^{-6}$ for the upper elements. A solution of such perturbed problem is presented in Fig. 4.2(b), we can see that in all the upper elements the material vanishes. At the same time $\tilde{Z}_h = Z_h = 1.75$, which means that both solutions $\mathscr{H}_h$ and $\tilde{\mathscr{H}}_h$ illustrated in Fig. 4.2(a) and (b), respectively, solve the unperturbed problem (FMD_h). This ambiguity may be considered an effect of combining the crude approximate equilibrium equations (cf. the definition of set $\Sigma_h(\Omega; F)$) with the very particular form of optimality condition (i) in Theorem 4.47: $\langle \epsilon_i, \sigma_i \rangle = \rho^0(\sigma_i)$. For a given ϵ_i this condition

allows to choose the magnitude of σ_i freely when $\rho(\epsilon_i) = 1$ while it enforces $\sigma_i = 0$ if $\rho(\epsilon_i) < 1$. In more complex problems than the uni-axial test here considered the situation is less ideal as it is more difficult to obtain $\rho(\epsilon_i) = 1$ in all elements when working in the finite dimensional space $\mathcal{U}^h(\Omega)$. It is then typical that in the optimal FE solution we observe regions of Ω where material is present in the upper simplices only and, at the same time, regions where the opposite is true thus exposing the full checker-board effect.

In [4] the checker-board issue is handled rather simply: for a quad mesh, in the post-processing the material properties are averaged over groups of four adjacent elements. In our setting, i.e. for triangular mesh, it will be enough to average over pairs of lower and upper simplices, the resulting material distribution is demonstrated in Fig. 4.2(c). Such action is justified in perspective of Conjecture 4.51 in which we expect that the approximate constitutive fields $\mathscr{H}_h \mathcal{L}^2 \llcorner \Omega$ converge weakly-* as measures only. Hence, we agree that all the figures will henceforward present the post-processed FE solution where $\mathscr{H}_h$ is averaged over quads, i.e. the unions of lower and upper simplices.

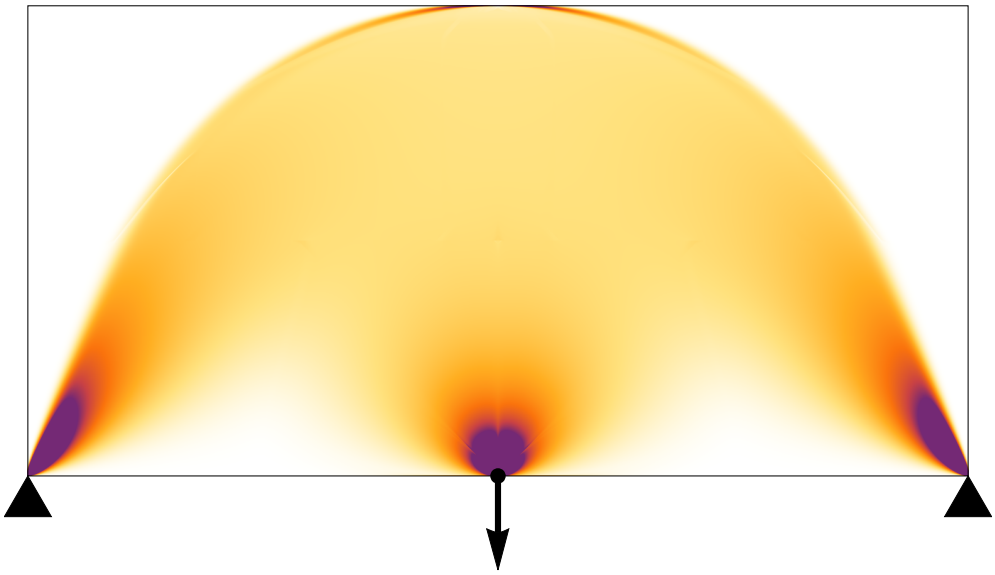


Figure 4.4: The bridge problem in the Anisotropic Material Design setting – distribution of $\mathrm{Tr}\mathscr{H}_h$ optimal in (FMD_h), heat-map of $\min\{\mathrm{Tr}\mathscr{H}_h, \alpha \|\mathrm{Tr}\mathscr{H}_h\|_{L^\infty}\}$ for $\alpha = 0.008$.

Employing the FE method developed two standard optimal design problems were engaged: the *cantilever problem* and the *bridge problem*; in this subsection we stay within the AMD setting. In both problems $\Omega = (-1, 1) \times (0, 1)$. In the first problem the body is supported on the whole left edge of the boundary, i.e. $\Upsilon_0 = [(-1, 0), (-1, 1)]$ while the load is a downward point-force applied at the centre of the right edge: $F = -\mathrm{e}_2\, \delta_{(1,1/2)}$. In the bridge problem we seek a structure that carries a point-force at the centre of the lower edge to *pin supports* at the bottom corners, namely $\Upsilon_0 = \{(-1, 0), (1, 0)\}$ and $F = -\mathrm{e}_2\, \delta_{(0,0)}$. Two mesh sizes were condsidered: a mesh spanned by 801×401 nodes counting 640 000 elements and a 1501×751 mesh with 2 250 000 elements. The computations were performed on a mobile workstation with Intel Core i7-6700HQ processor, 16 GB RAM and running

64-bit Windows 7. The graphical results, rendered in Mathematica® 11.1, are displayed in Fig. 4.3(a) (the cantilever problem) and Fig. 4.4 (the bridge problem) for the finer mesh only. The Dirichlet set Υ_0 is denoted by a hashed line in Fig. 4.4 and two triangular pin supports in Fig. 4.4. Since the constitutive fields $\mathscr{H}_h\mathcal{L}^2 \llcorner \Omega$ are expected to converge weakly-* to λ as measures it is natural that $\|\mathscr{H}_h\|_{L^\infty}$ may blow up as h tends to zero – e.g. when approximating an L^1 function with a singularity around some point or a lower dimensional part of λ, i.e. bar-like stiffeners. For this reason all the heat-maps showed in this section present a clipped function: $\min\{\mathrm{Tr}\mathscr{H}_h, \alpha\,\|\mathrm{Tr}\mathscr{H}_h\|_{L^\infty}\}$ for some small $\alpha > 0$; Fig. 4.3(b), (c) show the maps for higher and lower values of α for comparison.

In the subsequent subsections we will reformulate problems $(\mathbf{P}_h)$, $(\mathbf{P}_h^*)$ as conic quadratic programs for the next three settings of the Free Material Design studied in Section 3.3: IMD, FibMD and FibMD$_\pm$, the latter will be solved for $\kappa_+ = 1$, $\kappa_- = 1/3$. The summary of computational details for the cantilever and bridge problem may be found in Tables 4.1 and 4.2 respectively.

Remark 4.52. In Fig. 4.3(a) two peculiar singularities may be observed – they are positioned symmetrically with respect to the horizontal axis of symmetry, their coordinates are more or less $(0.6, 0.15)$ and $(0.6, 0.85)$. They form lower material density zones that are more or less 1D and they visually resemble cracks. Similar oddities are to be discerned in Fig. 4.4, see the vicinity of the point force and the two points symmetrically located on the edge of the optimal body. Such inclusions will reoccur in further simulations. Since the FEM herein developed is rather at an experimental stage, so far the present author did not investigate these issues to greater extent. It is expected that they are conditioned by the geometry of the mesh – they are often inclined under the 45° angle. Disturbing as they are they seem to vanish with h approaching zero. In the case of the IMD setting even stranger forms will emerge in the shape of ripple effects, see Fig. 4.6. Those, however, seem very regular therefore there might be a deeper explanation behind them.

Table 4.1: The cantilever problem: summary of FE computations for different anisotropy settings.

Anisotropy setting	Mesh (no. of elements)	Z_h	CPU time
AMD	640 000	6.3294	5 min 28 s
	2 250 000	6.3346	21 min 8 s
IMD	640 000	10.525	12 min 31 s
	2 250 000	10.534	47 min 36 s
FibMD	640 000	7.0000	7 min 12 s
($Z \approx 7.02471$, $Z_{\mathrm{X}^h} = 7.0275$)	2 250 000	7.0097	28 min 34 s
FibMD$_\pm$ for $\kappa_+ = 1$, $\kappa_- = \frac{1}{3}$	640 000	13.884	8 min 31 s
($Z \approx 13.936$, $Z_{\mathrm{X}^h} = 13.942$)	2 250 000	13.905	36 min 15 s

Table 4.2: The bridge problem: summary of FE computations for different anisotropy settings.

Anisotropy setting	Mesh (no. of elements)	Z_h	CPU time
AMD	640 000	2.3492	6 min 16 s
	2 250 000	2.3540	29 min 56 s
IMD	640 000	3.9083	11 min 34 s
	2 250 000	3.9177	64 min 15 s
FibMD	640 000	2.5671	8 min 55 s
($Z \approx 2.5708,\ Z_{\mathrm{X}^h} = 2.5721$)	2 250 000	2.5686	46 min 18 s
FibMD$_\pm$ for $\kappa_+ = 1$, $\kappa_- = \frac{1}{3}$	640 000	5.9121	8 min 27 s
($Z \approx 5.9208,\ Z_{\mathrm{X}^h} = 5.9235$)	2 250 000	5.9157	58 min 28 s

4.6.2.2. Conic quadratic formulation for the Isotropic Material Design setting

As already mentioned, for the IMD setting the cone $\mathcal{K}^4_\rho$ is not a quadratic cone nor any other type of cone that is available in the MOSEK toolbox. Indeed, we recall that for $d = 2$

$$\rho(\xi) = \max\left\{\frac{|\mathrm{Tr}\,\xi|}{\sqrt{2}}, \frac{|\mathrm{dev}\,\xi|}{\sqrt{2}}\right\}.$$

In what follows we shall rewrite the program $(\mathbf{P}_h)$, $(\mathbf{P}^*_h)$ so that the IMD anisotropy setting is captured using conic quadratic constraints only. We start by modifying the geometric matrices:

$$\mathbf{T}_k := \frac{1}{\sqrt{2}}\big(\mathbf{B}^{11}_k + \mathbf{B}^{22}_k\big), \qquad \mathbf{D}_k := \frac{1}{\sqrt{2}}\big(\mathbf{B}^{11}_k - \mathbf{B}^{22}_k\big).$$

We also redefine the isomorphism $\widetilde{\chi} : \mathbb{R}^3 \to \mathcal{S}^{2\times 2}$

$$\widetilde{\chi}(t, d, a^{12}) := t\,\frac{\mathrm{e}_1 \otimes \mathrm{e}_1 + \mathrm{e}_2 \otimes \mathrm{e}_2}{\sqrt{2}} + d\,\frac{\mathrm{e}_1 \otimes \mathrm{e}_1 - \mathrm{e}_2 \otimes \mathrm{e}_2}{\sqrt{2}} + a^{12}\sqrt{2}\,\mathrm{e}_1 \overset{s}{\otimes} \mathrm{e}_2. \tag{4.55}$$

It may be checked that, similarly to χ, the mapping $\widetilde{\chi}$ is an isometry as well. We discover that

$$\mathrm{Tr}\big(\widetilde{\chi}(t, d, a^{12})\big) = \sqrt{2}\,t, \qquad \big|\mathrm{dev}\big(\widetilde{\chi}(t, d, a^{12})\big)\big| = \sqrt{d^2 + (a^{12})^2}$$

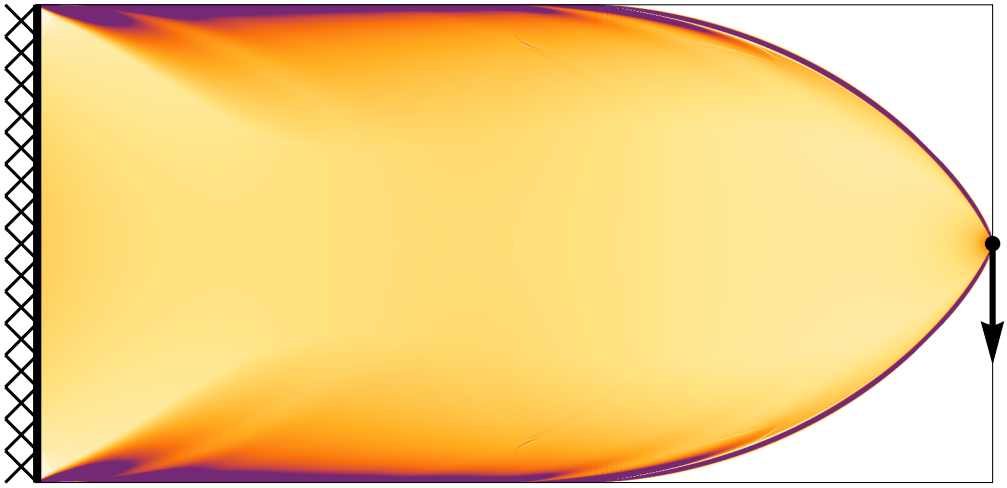
furnishing an equivalence

$$\rho\big(\widetilde{\chi}(t, d, a^{12})\big) \le 1 \qquad \Leftrightarrow \qquad -1 \le t \le 1 \quad \text{and} \quad \sqrt{d^2 + (a^{12})^2} \le \sqrt{2}.$$

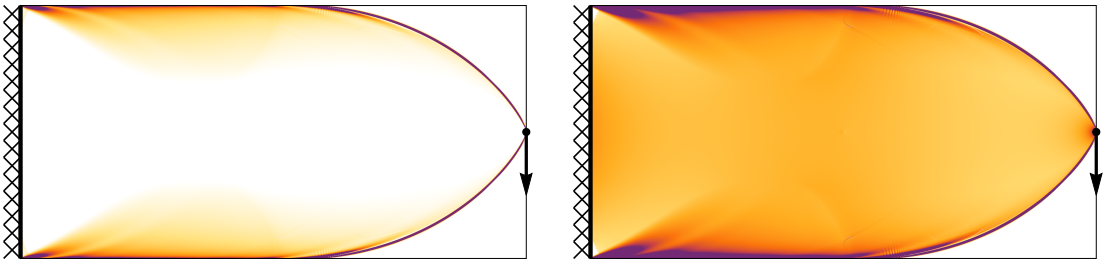
We are ready to write down a conic quadratic reformulation of $\mathbf{P}_h$:

$$Z_h = \max_{\substack{\mathbf{u}_1,\,\mathbf{u}_2 \in \mathbb{R}^m \\ \mathbf{t},\,\mathbf{d},\,\boldsymbol{\epsilon}^{12} \in \mathbb{R}^n, \\ \mathbf{r} \in \mathbb{R}^n}} \left\{ \langle \mathbf{u}_1, \mathbf{F}_1\rangle + \langle \mathbf{u}_2, \mathbf{F}_2\rangle \ : \ \begin{array}{c} \mathbf{T}_1\mathbf{u}_1 + \mathbf{T}_2\mathbf{u}_2 - \mathbf{t} = \mathbf{0} \\ \mathbf{D}_1\mathbf{u}_1 + \mathbf{D}_2\mathbf{u}_2 - \mathbf{d} = \mathbf{0} \\ \mathbf{B}^{12}_1\mathbf{u}_1 + \mathbf{B}^{12}_2\mathbf{u}_2 - \boldsymbol{\epsilon}^{12} = \mathbf{0} \\ \mathbf{r} = \sqrt{2}\,\mathbf{1} \\ -\mathbf{1} \le \mathbf{t} \le \mathbf{1} \\ \big(\mathbf{r}(i), \mathbf{d}(i), \boldsymbol{\epsilon}^{12}(i)\big) \in \mathcal{K}^3_{\mathrm{quad}} \qquad \forall\, i \end{array} \right\}.$$

The constraint $-\mathbf{1} \le \mathbf{t} \le \mathbf{1}$ ought to be understood component-wise, i.e. $-1 \le \mathbf{t}(i) \le 1$ for every i. This type of constraint is naturally available in the MOSEK software while in



(a)



(b) (c)

Figure 4.5: The cantilever problem in the Isotropic Material Design setting – distribution of: (a) $\mathrm{Tr}\mathscr{H}_h$, (b) bulk modulus $\mathscr{K}_h$, (c) shear modulus $\mathscr{G}_h$ optimal in (FMD_h).

the general conic programming framework from Section 2.6 may be recast by combining two conic and two equality constraints.

Skipping the derivation we give the problem that is dual to the above:

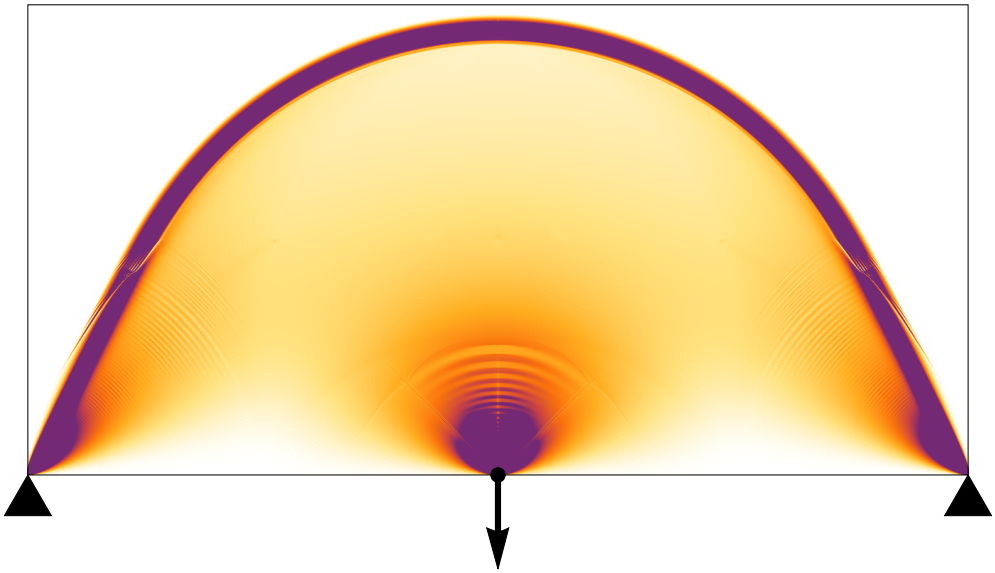
$$\min_{\substack{\mathbf{t}_+^*,\mathbf{t}_-^*\in\mathbb{R}^n_+\\ \mathbf{d}^*,\boldsymbol{\sigma}^{12}\in\mathbb{R}^n,\\ \mathbf{r}^0\in\mathbb{R}^n}} \left\{ \left\langle \mathbf{1}, \mathbf{t}_+^* + \mathbf{t}_-^* \right\rangle + \sqrt{2}\left\langle \mathbf{1}, \mathbf{r}^0 \right\rangle \;:\; \begin{array}{r} \mathbf{T}_1^\top(\mathbf{t}_+^* - \mathbf{t}_-^*) + \mathbf{D}_1^\top \mathbf{d}^* + (\mathbf{B}_1^{12})^\top \boldsymbol{\sigma}^{12} = \mathbf{F}_1 \\ \mathbf{T}_2^\top(\mathbf{t}_+^* - \mathbf{t}_-^*) + \mathbf{D}_2^\top \mathbf{d}^* + (\mathbf{B}_2^{12})^\top \boldsymbol{\sigma}^{12} = \mathbf{F}_2 \\ \left(\mathbf{r}^0(i), \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\right) \in \mathcal{K}^3_{\mathrm{quad}} \qquad \forall\, i \end{array} \right\}$$

By writing down the optimality conditions one may prove that necessarily $\mathbf{t}_+^*(i)\,\mathbf{t}_-^*(i) = 0$ at optimality. Based on solution of the problem above we recover solution σ_h of problem $(\mathcal{P}_h^*)$ via the following formula:

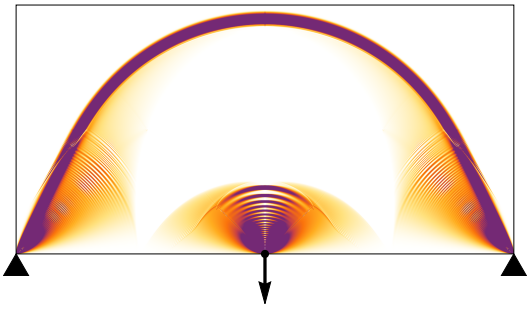
$$\sigma_h = \frac{1}{|E_i^h|}\,\widetilde{\chi}\Big(\mathbf{t}_+^*(i) - \mathbf{t}_-^*(i), \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\Big) \qquad \text{a.e. in } E_i^h;$$

note that $\mathbf{t}_+^*(i), \mathbf{t}_-^*(i) \geq 0$. Next, by acknowledging characterization (3.23) of the set $\underline{\mathcal{K}}_1(\sigma)$ we obtain optimal distributions $\mathscr{K}_h, \mathscr{G}_h \in L^1(\Omega;\mathbb{R}_+)$ of bulk and shear moduli respectively:

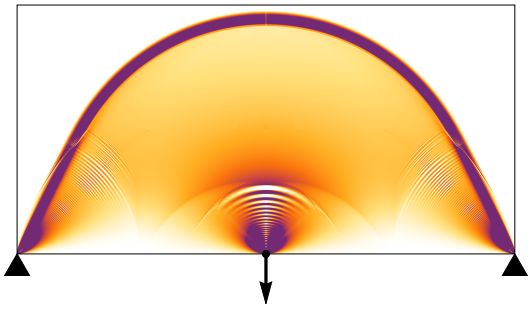
$$\mathscr{K}_h = \frac{1}{|E_i^h|}\,\frac{C_0}{Z_h}\,\frac{1}{2}\,\big(\mathbf{t}_+^*(i) + \mathbf{t}_-^*(i)\big), \qquad \mathscr{G}_h = \frac{1}{|E_i^h|}\,\frac{C_0}{Z_h}\,\frac{1}{2\sqrt{2}}\,\mathbf{r}^0(i) \qquad \text{a.e. in } E_i^h$$



(a)



(b)



(c)

Figure 4.6: The bridge problem in the Isotropic Material Design setting – distribution of: (a) $\mathrm{Tr}\mathscr{H}_h$, (b) bulk modulus $\mathscr{K}_h$, (c) shear modulus $\mathscr{G}_h$ that are optimal in (FMD_h).

which yield the optimal Hooke tensor function $\mathscr{H}_h \in L^1\big(\Omega; \mathscr{L}_+(\mathcal{S}^{2\times 2})\big)$ as below:

$$\mathscr{H}_h = 2\mathscr{K}_h \left(\frac{1}{2}\,\mathrm{I}\otimes\mathrm{I}\right) + 2\mathscr{G}_h \left(\mathrm{Id} - \frac{1}{2}\,\mathrm{I}\otimes\mathrm{I}\right);$$

we note that there holds

$$\mathrm{Tr}\mathscr{H}_h = 2\mathscr{K}_h + 4\mathscr{G}_h = \frac{1}{|E_i^h|}\,\frac{C_0}{Z_h}\Big(\mathbf{t}_+^*(i) + \mathbf{t}_-^*(i) + \sqrt{2}\,\mathbf{r}^0(i)\Big) = \frac{C_0}{Z_h}\,\rho^0(\sigma_h) \qquad \text{a.e. in } E_i^h.$$

The cantilever and the bridge problem were solved in the IMD setting; in both cases two mesh sizes were considered: counting 640 000 and 2 250 000 finite elements. For the finer mesh the heat-maps of functions $\mathrm{Tr}\mathscr{H}_h$, $\mathscr{K}_h$ and $\mathscr{G}_h$ that are optimal for the cantilever problem are showed in Fig. 4.5(a), (b) and (c) respectively, all the functions are clipped for $\alpha = 0.002$. For the bridge problem analogous results are presented in Fig. 4.6 where $\alpha = 0.005$ was chosen. For both problems one may observe that within the optimal body there emerge subregions where the bulk modulus vanishes. Computational details are given in Tables 4.1 and 4.2.

4.6.2.3. Conic quadratic formulation for the Fibrous Material Design setting

We recall that the d-dimensional (FMD) problem in the Fibrous Material Design setting is characterized by function ρ as bellow:

$$\rho(\xi) = \rho_{\pm}(\xi) := \max_{j\in\{1,\dots,d\}} \Big\{ \max\big\{\kappa_+\lambda_j(\xi), -\kappa_-\lambda_j(\xi)\big\} \Big\}$$

where κ_+, κ_- are positive numbers. In Examples 3.18 and 3.32 we have distinguished two cases: the (FibMD) setting where $\kappa_+ = \kappa_- = 1$ and $j_\pm = j$ induced the linear Hooke law and the (FibMD$_\pm$) setting where $\kappa_+ \neq \kappa_-$ for which $j_\pm$ induces a constitutive law of material that behaves differently under tension and compression. For matrix $\xi \in \mathcal{S}^{d\times d}$ the condition $\rho_\pm(\xi) \le 1$ is equivalent to bounding each of d eigenvalues $\lambda_j(\xi)$ by $-1/\kappa_-$ from below and $1/\kappa_+$ from above.

We will now focus on the 2D case and reformulate problems $(\mathbf{P}_h)$, $(\mathbf{P}_h^*)$ to a conic quadratic program. By utilizing the exact formulas for eigenvalues of a 2×2 matrix we arrive at equivalence below:

$$\forall\, j \quad -\frac{1}{\kappa_-} \le \lambda_j\big(\widetilde{\chi}(t,d,\epsilon^{12})\big) \le \frac{1}{\kappa_+} \qquad \Leftrightarrow \qquad -\frac{1}{\kappa_-} \le \frac{1}{\sqrt{2}}\, t \pm \frac{1}{\sqrt{2}}\sqrt{d^2 + (\epsilon^{12})^2} \le \frac{1}{\kappa_+}$$

where $\widetilde{\chi}$ is defined in (4.55). This paves the way to a conic quadratic representation of constraint $\rho_\pm\big(\widetilde{\chi}(t,d,\epsilon^{12})\big) \le 1$. As a result, the problem $(\mathbf{P}_h)$ gets to be rewritten as:

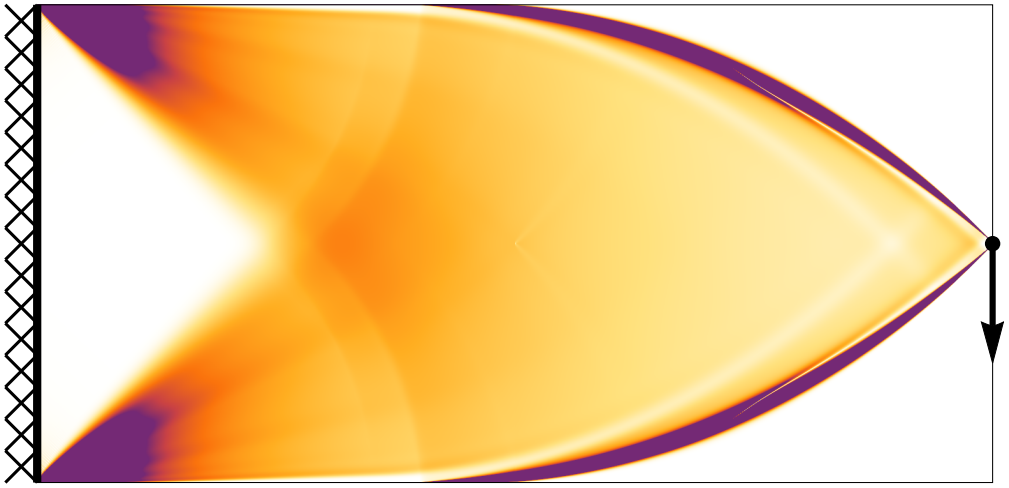
$$Z_h = \max_{\substack{\mathbf{u}_1,\mathbf{u}_2\in\mathbb{R}^m \\ \mathbf{t},\mathbf{d},\boldsymbol{\epsilon}^{12}\in\mathbb{R}^n, \\ \mathbf{r}\in\mathbb{R}^n}} \left\{ \langle \mathbf{u}_1, \mathbf{F}_1\rangle + \langle \mathbf{u}_2, \mathbf{F}_2\rangle \ : \ \begin{array}{c} \mathbf{T}_1\mathbf{u}_1 + \mathbf{T}_2\mathbf{u}_2 - \mathbf{t} = \mathbf{0} \\ \mathbf{D}_1\mathbf{u}_1 + \mathbf{D}_2\mathbf{u}_2 - \mathbf{d} = \mathbf{0} \\ \mathbf{B}_1^{12}\mathbf{u}_1 + \mathbf{B}_2^{12}\mathbf{u}_2 - \boldsymbol{\epsilon}^{12} = \mathbf{0} \\ \mathbf{t} + \mathbf{r} \le \frac{\sqrt{2}}{\kappa_+}\mathbf{1} \\ -\frac{\sqrt{2}}{\kappa_-}\mathbf{1} \le \mathbf{t} - \mathbf{r} \\ \big(\mathbf{r}(i), \mathbf{d}(i), \boldsymbol{\epsilon}^{12}(i)\big) \in \mathcal{K}^3_{\rm quad} \qquad \forall\, i \end{array} \right\}$$

and the dual problem reads

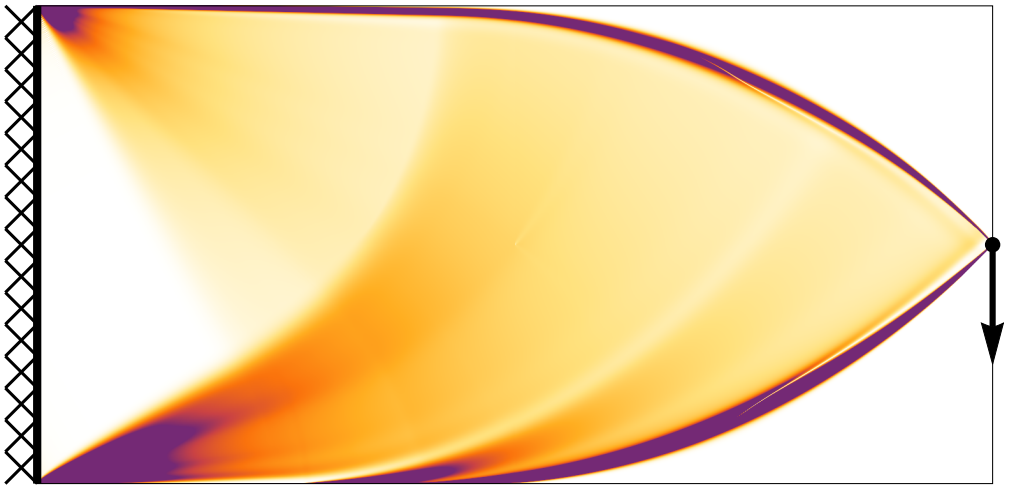
$$\min_{\substack{\boldsymbol{\lambda}_+,\boldsymbol{\lambda}_-\in\mathbb{R}^n_+ \\ \mathbf{d}^*,\boldsymbol{\sigma}^{12}\in\mathbb{R}^n, \\ \mathbf{r}^0\in\mathbb{R}^n}} \left\{ \frac{\sqrt{2}}{\kappa_+}\langle \mathbf{1}, \boldsymbol{\lambda}_+\rangle + \frac{\sqrt{2}}{\kappa_-}\langle \mathbf{1}, \boldsymbol{\lambda}_-\rangle \ : \ \begin{array}{c} \mathbf{T}_1^\top(\boldsymbol{\lambda}_+ - \boldsymbol{\lambda}_-) + \mathbf{D}_1^\top\mathbf{d}^* + (\mathbf{B}_1^{12})^\top\boldsymbol{\sigma}^{12} = \mathbf{F}_1 \\ \mathbf{T}_2^\top(\boldsymbol{\lambda}_+ - \boldsymbol{\lambda}_-) + \mathbf{D}_2^\top\mathbf{d}^* + (\mathbf{B}_2^{12})^\top\boldsymbol{\sigma}^{12} = \mathbf{F}_2 \\ \boldsymbol{\lambda}_+ + \boldsymbol{\lambda}_- - \mathbf{r}^0 \ge \mathbf{0} \\ \big(\mathbf{r}^0(i), \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\big) \in \mathcal{K}^3_{\rm quad} \qquad \forall\, i \end{array} \right\}$$

In order to see how the variables $\boldsymbol{\lambda}_+(i), \boldsymbol{\lambda}_-(i) \ge 0$ are linked with the eigenvalues of the matrix $\widetilde{\chi}\big(\boldsymbol{\lambda}_+(i) - \boldsymbol{\lambda}_-(i), \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\big)$ one may check that there holds:

$$\left\{ \begin{array}{l} \sigma = \widetilde{\chi}(\lambda_+ - \lambda_-, d^*, \sigma^{12}), \\ \lambda_+, \lambda_- \in \mathbb{R}_+, \\ \lambda_+ + \lambda_- \ge \sqrt{(d^*)^2 + (\sigma^{12})^2} \end{array} \right. \qquad \Rightarrow \qquad \left\{ \begin{array}{l} \sqrt{2}\,\lambda_+ = \sum_{j=1}^2 \max\{\lambda_j(\sigma), 0\}, \\ \sqrt{2}\,\lambda_- = \sum_{j=1}^2 \big(-\min\{\lambda_j(\sigma), 0\}\big), \end{array} \right.$$



(a)



(b)

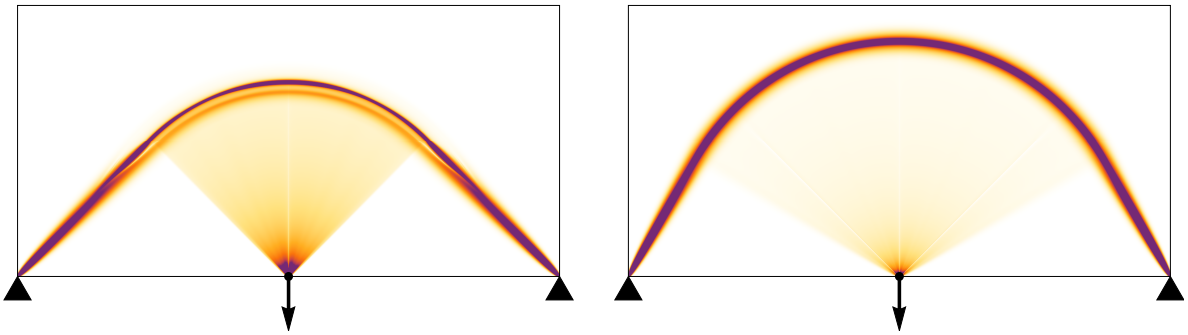
Figure 4.7: The cantilever problem in the Fibrous Material Design setting – distribution of $\mathrm{Tr}\mathscr{H}_h$: (a) the FibMD case where $\kappa_+ = \kappa_- = 1$; (b) the FibMD$_\pm$ case where $\kappa_+ = 1$, $\kappa_- = 1/3$.

which allows to recognize the objective function in the minimization problem above as $\rho^0_\pm\left(\widetilde{\chi}\big(\boldsymbol{\lambda}_+(i) - \boldsymbol{\lambda}_-(i), \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\big)\right)$, cf. Example 3.32. For solution of this minimization problem the stress field $\sigma_h \in L^1(\Omega; \mathcal{S}^{2\times 2})$ satisfying

$$\sigma_h = \frac{1}{|E_i^h|}\,\widetilde{\chi}\Big(\boldsymbol{\lambda}_+(i) - \boldsymbol{\lambda}_-(i), \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\Big) \qquad \text{a.e. in } E_i^h$$

solves the problem $(\mathcal{P}_h^*)$. By combining (4.54) with the formula for optimal Hooke tensor $\underline{H}_\sigma$ in Example 3.32 we obtain the optimal Hooke tensor function for the Fibrous Material Design problem:

$$\mathscr{H}_h = \frac{C_0}{Z_h}\sum_{j=1}^{2}\max\left\{\frac{1}{\kappa_+}\lambda_j(\sigma_h), -\frac{1}{\kappa_-}\lambda_j(\sigma_h)\right\}\; v_j(\sigma_h)\otimes v_j(\sigma_h)\otimes v_j(\sigma_h)\otimes v_j(\sigma_h)$$



(a) (b)

Figure 4.8: The bridge problem in the Fibrous Material Design setting – distribution of $\mathrm{Tr}\mathscr{H}$: (a) the FibMD case where $\kappa_+ = \kappa_- = 1$; (b) the FibMD$_\pm$ case where $\kappa_+ = 1$, $\kappa_- = 1/3$.

where $v_j\big(\sigma_h(x)\big)$ is the unit eigenvector of $\sigma_h(x)$ corresponding to eigenvalue $\lambda_j\big(\sigma_h(x)\big)$. Point-wise we compute the trace of the Hooke tensor:

$$\begin{aligned}\mathrm{Tr}\mathscr{H}_h &= \frac{C_0}{Z_h}\left(\frac{1}{\kappa_+}\left(\sum_{j=1}^{2}\max\{\lambda_j(\sigma_h),0\}\right) - \frac{1}{\kappa_-}\left(\sum_{j=1}^{2}\min\{\lambda_j(\sigma_h),0\}\right)\right)\\ &= \frac{C_0}{Z_h}\frac{1}{|E_i^h|}\left(\frac{\sqrt{2}}{\kappa_+}\boldsymbol{\lambda}_+(i) + \frac{\sqrt{2}}{\kappa_-}\boldsymbol{\lambda}_-(i)\right) \qquad \text{a.e. in } E_i^h.\end{aligned}$$

Both the cantilever and the bridge problem were solved for the Fibrous Material Design setting in two cases: the case (FibMD) where $\kappa_+ = \kappa_- = 1$ and the case (FibMD$_\pm$) for $\kappa_+ = 1$, $\kappa_- = 1/3$ in which tension is favoured over compression. Computations were performed for two mesh sizes: counting 640 000 and 2 250 000 elements, the details may be found in Tables 4.1 and 4.2. The heat-maps of optimal function $\mathrm{Tr}\mathscr{H}_h$ are showed in Fig. 4.7 (clipped for $\alpha = 0.002$) and Fig. 4.8 (clipped for $\alpha = 0.015$) for the cantilever and the bridge problem respectively.

4.7. Michell problem and related topics

In this section we review the renown Michell problem of designing a fully stressed framework structure of the least volume. The problem was proposed by A.G.M. Michell at the dawn of the 20th century (cf. [127]) as a relaxation of the truss optimization problem. The Michell's theory was developed in a number of works, [105, 162, 114, 116] to name a few. An extensive coverage on the theoretical and numerical aspects of Michell problem may be found in [118]. In the paper [42] the Michell problem was for the first time posed in the measure-theoretic setting, i.e. as the pair of type $(\mathcal{P})$ and $(\mathcal{P}^*)$.

Our motivation for visiting Michell problem is twofold. The first premise is simple enough: it will appear that Michell problem is equivalent to the (FMD) problem in the Fibrous Material Design setting that was herein proposed. The second and more important reason is to prepare the background for subsequent topics of this dissertation. In Chapter

5 we will encounter problems that share some mathematical similarities with Michell's theory, while Chapter 6 puts forward a new design problem that draws direct inspiration from Michell problem.

We will start by discussing equilibrium of trusses that may be treated as a point of departure towards Michell structures.

4.7.1. Measure-theoretic approach to truss structures

Truss is a structure composed of finite number of straight bars whose ends meet at joints where a point load is applied; the load is transferred between the joints by means of axial forces: tensile or compressive. Classically, one models each bar individually as a 1D body and then make the synthesis by writing down kinematic compatibility conditions and equilibrium equations at the joints. Instead, we shall adapt an approach proposed in [42] where truss is *a priori* a d-dimensional body represented by a measure supported on a graph, i.e. a union of segments. Consequently, truss as an elastic body can be given through a constitutive field $\lambda = \mathscr{H}\mu$ where $\mu = \mathcal{H}^1 \llcorner \mathcal{G}$ and $\mathcal{G} \subset \mathbb{R}^d$ is a graph, $\mathscr{H}$ can for instance be a constant isotropic Hooke tensor. The load $F \in \mathcal{M}^b(\mathbb{R}^d; \mathbb{R}^d)$ supported on the set of joints induces the relative stress σ that satisfies equilibrium equation $-\mathrm{Div}(\sigma\mu) = F$. Since μ is a multi-junction measure (cf. Definition 2.58) from Theorem 2.57 we infer that σ must be rank-one, i.e. $\sigma = S\,\tau \otimes \tau$ where $S \in \mathbb{R}$ is the axial force varying from bar to bar. This way we recover the classical description of trusses. Below we concentrate on the resulting absolute stress $\varsigma = \sigma\mu$ and the equilibrium equation $-\mathrm{Div}\,\varsigma = F$ only. We shall assume that the truss is fixed on the Dirichlet region Υ_0 being a compact subset of $\mathbb{R}^d$ therefore the equation will be understood in the sense of distributions in $\mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$.

Let us agree that by $\Delta := \big\{(x,x) : x \in \mathbb{R}^d\big\}$ we will henceforward denote the diagonal of the Cartesian product $\mathbb{R}^d \times \mathbb{R}^d$. For a pair of distinct points $(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \setminus \Delta$ we define the absolute stress in a bar connecting x to y and subject to unit tensile axial force:

$$\varsigma^{x,y} := \tau^{x,y} \otimes \tau^{x,y}\ \mathcal{H}^1 \llcorner [x,y] \qquad \text{where} \quad \tau^{x,y} := \frac{y-x}{|y-x|}.$$

We shall now compute the divergence $-\mathrm{Div}\,\varsigma^{x,y}$ in $\mathbb{R}^d \setminus \Upsilon_0$. Let $s : [0, |y-x|] \to [x,y]$ be the natural parametrization of the closed segment $[x,y]$, i.e. $s(t) = x + t\,\tau^{x,y}$. For any test function $\phi \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$ we define $\bar{\phi}(t) := \big\langle \phi\big(s(t)\big), \tau^{x,y} \big\rangle$ to find that

$$\int \langle \nabla\phi, \varsigma^{xy} \rangle = \int_{[x,y]} \langle \nabla\phi, \tau^{x,y} \otimes \tau^{x,y} \rangle\, d\mathcal{H}^1 = \int_0^{|y-x|} \bar{\phi}'(t)\, dt = \langle \phi(y), \tau^{x,y} \rangle - \langle \phi(x), \tau^{x,y} \rangle. \tag{4.56}$$

As a consequence

$$-\mathrm{Div}\,\varsigma^{x,y} = \mathbb{1}_{\mathbb{R}^d \setminus \Upsilon_0}(y)\,\tau^{x,y}\,\delta_y - \mathbb{1}_{\mathbb{R}^d \setminus \Upsilon_0}(x)\,\tau^{x,y}\,\delta_x \qquad \text{in } \ \mathbb{R}^d \setminus \Upsilon_0$$

where $\mathbb{1}_B$ is the characteristic function of set B, i.e. $\mathbb{1}_B(x) = 1$ for $x \in B$ and $\mathbb{1}_B(x) = 0$ for $x \notin B$. We have thus recovered a standard fact from mechanics of structures: if $[x,y] \cap \Upsilon_0 = \varnothing$, to exert a unit tensile axial force we have to apply two unit forces at both ends of the bar, in directions that are tangent to that bar and in the outward fashion.

From one bar we move on to a truss. Let us take a convex bounded domain $\Omega \subset \mathbb{R}^d$. For integer m let us assume a set of joints $\{A_i\}_{i=1}^m \subset \overline{\Omega}$ which "spans the truss" – every pair of distinct joints (A_i, A_j) is a potential bar of axial force $S_{ij} \in \mathbb{R}$ (tensile force if $S_{ij} > 0$ and compressive if $S_{ij} < 0$). Statics of truss may readily be described by the absolute stress $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d})$ and a load $F \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ applied at joints:

$$\varsigma = \sum_{\substack{i,j=1 \\ i\neq j}}^{m} S_{ij}\, \varsigma^{A_i,A_j}, \qquad F = \sum_{j=1}^{m} F_j\, \delta_{A_j} \tag{4.57}$$

where $F_j \in \mathbb{R}^d$. Note that the fact that $\varsigma^{A_i,A_j} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d})$ is by convexity of Ω. By acknowledging (4.56) we arrive at the algebraic representation of truss' equilibrium equations:

$$-\mathrm{Div}\,\varsigma = F \quad \text{in } \mathbb{R}^d \setminus \Upsilon_0 \qquad \Leftrightarrow \qquad \sum_{i=1,\ i\neq j}^{m} (S_{ij} + S_{ji})\, \tau^{A_i,A_j} = F_j \qquad \forall A_j \in \overline{\Omega} \setminus \Upsilon_0.$$

The vector algebraic equations on the RHS can be recognized as equilibrium equations of joints which are not fixed. We can see that the whole sum $S_{ij} + S_{ji}$ is effectively the axial force in a bar connecting A_i to A_j, which creates a slight inconvenience. To avoid ambiguity we may enforce symmetry $S_{ij} = S_{ji}$.

One of the main ideas of paper [42] was to generalize the notion of truss measures ς above to trusses composed of infinite number of bars. Inspired by the framework of optimal transport problem (cf. Section 4.8) to that aim we employ a pairing measure $\mathscr{S} \in \mathcal{M}^b(\overline{\Omega} \times \overline{\Omega} \setminus \Delta; \mathbb{R})$ being a finite Radon measure on the locally compact space. To it we can assign a measure $\varsigma_{\mathscr{S}} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d})$ that is defined as a linear functional via an integration formula – for any function $\epsilon \in C(\overline{\Omega}; \mathcal{S}^{d\times d})$

$$\begin{aligned} \langle \epsilon, \varsigma_{\mathscr{S}} \rangle_{(C,\mathcal{M})} &:= \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} \left(\int \langle \epsilon, \varsigma^{x,y} \rangle \right) \mathscr{S}(dxdy) \\ &= \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} \left(\int_{[x,y]} \langle \epsilon(s), \tau^{x,y} \otimes \tau^{x,y} \rangle\, \mathcal{H}^1(ds) \right) \mathscr{S}(dxdy). \end{aligned} \tag{4.58}$$

Indeed, the mapping $(x, y) \mapsto \int \langle \epsilon, \varsigma^{x,y} \rangle$ is continuous and thus Borel measurable and secondly, since $\mathscr{S}$ is of compact support, we have $\left| \int \langle \epsilon, \varsigma^{x,y} \rangle \right| \leq \|\epsilon\|_{\infty,\Omega}\, |y - x| \leq \|\epsilon\|_{\infty,\Omega}\, \mathrm{diam}(\Omega)$ therefore (x, y) is $\mathscr{S}$-measurable and $\mathscr{S}$-essentially bounded in the sense of Definition 2.26. One can see that the integral formula for $\varsigma_{\mathscr{S}}$ generalizes the absolute stress ς of a truss defined through double sum in (4.57) while measure $\mathscr{S}$ itself is an infinite dimensional counterpart of axial forces S_{ij}. Henceforward we shall refer to $\mathscr{S}$ as a *generalized truss*.

Remark 4.53. Note that we may put, by convention, that $\tau^{x,y} = 0$ for $x = y$ and then we can operate with $\mathscr{S} \in \mathcal{M}(\overline{\Omega} \times \overline{\Omega}; \mathbb{R})$ instead, which is often more convenient, see for instance Section 5.5.

As the next step we shall derive the analogue of truss' joints equilibrium equation given above in the finite case. For that purpose we require a linear *geometric operator* $\mathcal{B} : C_0(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d) \to C_0(\overline{\Omega} \times \overline{\Omega} \setminus \Delta; \mathbb{R})$ which in classical mechanics of trusses is given as

a finite dimensional rectangular matrix of the so called *directional cosines*. For any distinct pair of points (x, y) we define

$$(\mathcal{B}u)(x,y) := \Big\langle u(y) - u(x), \tau^{x,y} \Big\rangle = \Big\langle u(y) - u(x), \frac{y-x}{|y-x|} \Big\rangle. \tag{4.59}$$

Readily, for a vector displacement function u the number $(\mathcal{B}u)(x, y)$ informs of elongation of a bar that potentially connects x and y (or contraction if the number is negative). We are ready to examine equilibrium of a generalized truss:

Proposition 4.54 (**Equilibrium of generalized trusses**). *For a bounded convex domain $\Omega \subset \mathbb{R}^d$ the load F being any vector valued measure in $\mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ consider a generalized truss measure $\mathscr{S} \in \mathcal{M}(\overline{\Omega} \times \overline{\Omega} \setminus \Delta; \mathbb{R})$. Then, equilibrium of the truss stress $\varsigma_{\mathscr{S}}$*

$$-\mathrm{Div}\, \varsigma_{\mathscr{S}} = F \qquad \text{in} \quad \mathbb{R}^d \setminus \Upsilon_0 \tag{4.60}$$

is equivalent to the equilibrium of joints (virtually of each point in $\mathbb{R}^d \setminus \Upsilon_0$)

$$\mathcal{B}^* \mathscr{S} = F \qquad \text{in} \quad \mathcal{M}^b(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d). \tag{4.61}$$

Proof. By definition of $\varsigma_{\mathscr{S}}$ and of the distributional divergence, for any smooth function $\phi \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$ there holds

$$\langle \phi, -\mathrm{Div}\, \varsigma_{\mathscr{S}} \rangle_{(\mathcal{D},\mathcal{D}')} = \int \langle \nabla \phi, \varsigma_{\mathscr{S}} \rangle = \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} \left(\int_{[x,y]} \langle \nabla\phi, \tau^{x,y} \otimes \tau^{x,y} \rangle \, d\mathcal{H}^1 \right) \mathscr{S}(dxdy)$$

$$= \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} \langle \phi(y) - \phi(x), \tau^{x,y} \rangle \, \mathscr{S}(dxdy) = \langle \mathcal{B}\phi, \mathscr{S} \rangle_{(C_0,\mathcal{M}^b)} = \langle \phi, \mathcal{B}^* \mathscr{S} \rangle_{(C_0,\mathcal{M}^b)}$$

where we have utilised (4.56); note that the last two pairings are intended for functional spaces on $\overline{\Omega}\times\overline{\Omega}\setminus\Delta$ and $\mathbb{R}^d\setminus\Upsilon_0$, respectively. Implication (4.60) $\Leftarrow$ (4.61) is thus immediate whilst the converse follows by density of $\mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$ in $C_0(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$. □

4.7.2. The Fibrous Material Design problem is the Michell problem

To channel the truss-like absolute stress ς given in (4.57) a truss, in the form of material distribution that charges the graph spanned by joints (vertices) $\{A_i\}_{i=1}^m$, ought to be designed. Instead of considering elasticity, however, we may choose to assume that the material is perfectly plastic in which the relative stress (its normal component) must lie within bounds $[-\bar{s}_-, \bar{s}_+]$ where $\bar{s}_- > 0$ and $\bar{s}_+ > 0$ are the yield stresses for compression and tension respectively. The objective of the design process is to choose cross-sectional area of each bar $[A_i, A_j]$ so that truss' volume is minimal. If the axial force S_{ij} is prescribed, the yield criterion is met if the area is non-smaller than two numbers: $-S_{ij}/\bar{s}_-$ and $S_{ij}/\bar{s}_+$ (or simply $|S_{ij}|/\bar{s}$ when $\bar{s} = \bar{s}_- = \bar{s}_+$). Therefore, the minimal volume of a truss that channels the absolute stress ς can be computed as:

$$V(\varsigma) = \sum_{i,j=1}^m |A_i - A_j| \max\left\{ \frac{S_{ij}}{\bar{s}_+}, -\frac{S_{ij}}{\bar{s}_-} \right\}.$$

The *plastic truss design problem* may be thus defined as follows: for a convex bounded design domain $\Omega \subset \mathbb{R}^d$, a Dirchlet set $\Upsilon_0 \subset \overline{\Omega}$ and a load $F \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ of finite support we seek the set of joints $\{A_i\}_{i=1}^m \subset \overline{\Omega}$ and a truss stress $\varsigma = \sum_{i,j=1}^m S_{ij}\, \varsigma^{A_i,A_j}$ satisfying

equilibrium $-\mathrm{Div}\,\varsigma = F$ in $\mathbb{R}^d \setminus \Upsilon_0$ for which $V(\varsigma)$ is minimal. Since the work of Michell [127] it is well known that such problem is badly posed: in general the best we can do is to propose a sequence of trusses ς_h for which the number of bars increases to infinity while the volume $V(\varsigma_h)$ approaches its lower bound, cf. [118, Fig. 3.1 in Section 3.1] or [42, Fig. 3]. The idea of Michell was to relax the truss problem by allowing framework-type bodies that combine "continuous" medium and 1D bars, possibly curved.

The stress-based form of the Michell problem was for the first time posed in [162] while its rigorous, measure-theoretic setting was given in [42] (for $\bar{s}_+ = \bar{s}_- = 1$ and $\Upsilon_0 = \varnothing$):

$$\inf\left\{\int c_{\mathrm{M}}(\varsigma) \,:\, \varsigma \in \mathcal{M}\big(\overline{\Omega};\mathcal{S}^{d\times d}\big),\quad -\mathrm{Div}\,\varsigma = F \ \text{ in } \mathbb{R}^d\setminus\Upsilon_0\right\} \tag{MP}$$

where $c_{\mathrm{M}} : \mathcal{S}^{d\times d} \to \mathbb{R}_+$ is the Michell-type cost being a closed gauge defined through the eigenvalues:

$$c_{\mathrm{M}}(\sigma) = \sum_{i=1}^{d} \max\left\{\frac{\lambda_i(\sigma)}{\bar{s}_+}, -\frac{\lambda_i(\sigma)}{\bar{s}_-}\right\}. \tag{4.62}$$

The dual, displacement-based problem was derived by Michell himself:

$$\sup\left\{\int \langle u, F\rangle \ :\ u \in \mathcal{D}\big(\mathbb{R}^d\setminus\Upsilon_0;\mathbb{R}^d\big),\quad -\frac{1}{\bar{s}_-} \le \lambda_i\big(e(u)\big) \le \frac{1}{\bar{s}_+} \quad \text{in } \overline{\Omega}\ \ \forall\, i\right\} \tag{MP*}$$

The set of admissible functions u in (MP*) can be written as $\mathcal{U}_1$ provided we take $\rho = c^0_{\mathrm{M}}$, cf. (4.9). In [42] the authors observed that, once convexity of Ω is assumed, the point-wise bound on $\lambda_i\big(e(u)\big)$ can be reinterpreted as a two-point condition imposed on the elongations $\mathcal{B}u = (\mathcal{B}u)(x,y)$. By means of a subtle duality argument it allowed them to establish a new rigorous link to the plastic design of generalized truss $\mathscr{S}$:

Theorem 4.55 (**Bouchitté-Gangbo-Seppecher**). *Assuming that Ω is convex and $\Upsilon_0 = \varnothing$ let us consider Michell problem in form of the pair* (MP), (MP*) *for $\bar{s}_+ = \bar{s}_- = 1$. Then, the set $\overline{\mathcal{U}}_1$ of admissible functions u can be characterized via a two-point condition:*

$$\overline{\mathcal{U}}_1 = \Big\{u \in C(\overline{\Omega};\mathbb{R}^d) \,:\, -|y-x| \le (\mathcal{B}u)(x,y) \le |y-x| \quad \forall\,(x,y)\in\overline{\Omega}\times\overline{\Omega}\setminus\Delta\Big\}.$$

Moreover, assuming that $\int \langle u_0, F\rangle = 0$ for every $u_0 \in \mathcal{R}$

$$\max\left\{\int_{\overline{\Omega}} \langle u, F\rangle \,:\, u \in \overline{\mathcal{U}}_1\right\} \tag{4.63}$$

$$= \inf\left\{\int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} |y-x|\,|\mathscr{S}|(dxdy) \,:\, \mathscr{S}\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega}\setminus\Delta;\mathbb{R}),\ \mathcal{B}^*\mathscr{S} = F\right\} \tag{4.64}$$

$$= \inf\left\{\int_{\overline{\Omega}} c_{\mathrm{M}}\big(\varsigma_{\mathscr{S}}\big) \,:\, \mathscr{S}\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega}\setminus\Delta;\mathbb{R}),\ \mathcal{B}^*\mathscr{S} = F\right\} \tag{4.65}$$

$$= \min\left\{\int_{\overline{\Omega}} c_{\mathrm{M}}(\varsigma) :\ \varsigma \in \mathcal{M}\big(\overline{\Omega};\mathcal{S}^{d\times d}\big),\quad -\mathrm{Div}\,\varsigma = F \ \ \textit{in } \mathbb{R}^d\setminus\Upsilon_0\right\} \tag{4.66}$$

whereas the first and the last problem attains a solution.

With $\bar{s}_+ = \bar{s}_- = 1$ one readily recognizes the functional $\int |y-x|\,|\mathscr{S}|(dxdy)$ as the generalization of volume $V(\varsigma)$ towards infinite trusses $\mathscr{S}$. The optimal truss problem (4.64), however, does not have a solution in general – loosely speaking, the Michell structure

is composed of infinite number of cables, i.e. 1D elements that may be curved, rather than straight bars. In [42] the authors made an effort to extend the problem (4.64) to measures $\mathscr{S}$ on a set of regular curves. Ultimately, the existence of solutions in form of such measures was not proved in general, although all the analytical solutions ς of Michell problem (MP) available in the literature may indeed be disintegrated with respect to such set of curves.

Remark 4.56. Limitation of Theorem 4.55 to the case $\bar{s}_+ = \bar{s}_- = 1$ (and, in fact, to $\Upsilon_0 = \varnothing$ as well) is merely due to the particular scope of the work [42]. The present author expects that the results can be brought to the asymmetric setting by the following alterations: in the characterization of $\overline{\mathcal{U}}_1$ as the lower and upper bounds for $(\mathcal{B}u)(x,y)$ we put $-|y-x|/\bar{s}_-$ and $|y-x|/\bar{s}_+$, respectively, while in (4.64) we change the objective functional to the sum $\int |y-x|\big(\frac{1}{\bar{s}_+}\mathscr{S}_+ + \frac{1}{\bar{s}_-}\mathscr{S}_-\big)(dxdy)$ where $\mathscr{S}_+, \mathscr{S}_-$ is the Jordan's decomposition of $\mathscr{S}$.

With Michell problem formulated as the pair (MP), (MP*) above we can easily find a link to the Free Material Design problem in the framework of 2D/3D elasticity: we observe that the Michell cost function c_M is precisely the closed gauge $\rho^0_\pm$ that emerges in the Fibrous Material Design setting of the (FMD) problem if one chooses $\kappa_+ = \bar{s}_+$ and $\kappa_- = \bar{s}_-$, cf. Example 3.32. Aa a result, the problems (MP) and ($\mathcal{P}^*$) in the FibMD$\pm$ setting coincide and the same goes for problems (MP*) and ($\mathcal{P}$). In the FibMD$_\pm$ setting the formula for optimal Hooke tensor $\underline{H}_\sigma$ for a stress $\sigma \in \mathcal{S}^{d\times d}$ at a point (i.e. formula for $\underline{H}_\sigma \in \underline{\mathcal{K}}_1(\sigma)$) reads

$$\underline{H}_\sigma = \sum_{i=1}^{d} \frac{\max\big\{\frac{1}{\kappa_+}\lambda_i(\sigma), -\frac{1}{\kappa_-}\lambda_i(\sigma)\big\}}{\rho^0_\pm(\sigma)}\; v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma)$$

where $v_i(\sigma)$ is an eigenvector corresponding to $\lambda_i(\sigma)$, see (3.48). One might say that to realize $\underline{H}_\sigma$ at a point we must lay three orthogonal fibres along principal directions of stress σ and of cross-sectional areas proportional to respective principal values of σ. This is reminiscent of the concept of designing the distribution of plastic material in Michell structures. Putting this differently, if ς is a solution of (MP) (or equivalently ($\mathcal{P}^*$)) the constitutive field $\lambda = \frac{C_0}{Z}\,\mathscr{H}\mu$ where $\mu = \rho^0_\pm(\varsigma)$ and $\mathscr{H}(x) = \underline{H}_{\sigma(x)}$ for $\sigma = \frac{d\varsigma}{d\mu}$ (cf. Corollary 4.7) is optimal for the (FMD) problem in FibMD$_\pm$ setting and may be viewed as an elastic representative of Michell structure. In summary:

The (FMD) *problem in the setting of Fibrous Material Design with dissymmetry in tension and compression for parameters κ_+, κ_- is equivalent to Michell problem* (MP) *for limit stresses $\bar{s}_+$ and $\bar{s}_-$ in tension and, respectively, compression provided that $\kappa_+/\kappa_- = \bar{s}_+/\bar{s}_-$.*

The reason for this connection is to be sought in the particular form of the cone of admissible Hooke tensors that virtually defines the Fibrous Material Design setting: we recall that $\mathcal{K} = \mathcal{K}_{fib} := \mathrm{co}(\mathcal{K}_{axial})$ where $\mathcal{K}_{axial}$ is the non-convex cone of uni-axial Hooke tensors $H = a\,\tau\otimes\tau\otimes\tau\otimes\tau$ with $a \geq 0$, $\tau \in S^{d-1}$. For such Hooke tensors H at a point the material mimics the action of a bar as it produces stress $\sigma = H\xi = a\,\langle \xi, \tau\otimes\tau\rangle\,\tau\otimes\tau$

(in the case when $\kappa_+ = \kappa_-$), cf. the form of the energy j_μ in (4.29) and the comment below it. For the well-posedness of the optimization problem the convex hull $\mathrm{co}(\mathcal{K}_{axial})$ must be taken thus arriving at design of material composed of fibres running in arbitrary directions, which ultimately turns out to be equivalent do designing a Michell structure.

Another works where a link between the Michell problem and optimal design of elastic body was made are [5], [15], [46], [34] where in the case when $\bar{s}_+ = \bar{s}_-$ Michell problem was recovered as the asymptotic limit for structural topology design problem in the high-porosity regime. To the present author's knowledge, the (FMD) problem in the $\mathrm{FibMD}_\pm$ setting is the first formulation of elastic design problem that is equivalent to Michell problem for non-equal permissible stresses $\bar{s}_+ \neq \bar{s}_-$.

4.7.3. An alternative numerical scheme for Michell problem – the ground structure approach

With the established relation between Michell problem and the (FMD) problem in the $\mathrm{FibMD}_\pm$ setting we find that the finite element method put forward in Section 4.6.2.3 is in fact a numerical scheme for Michell problem. As a result the numerical solutions of the cantilever and the bridge problems showed in Fig. 4.7 and, respectively, Fig. 4.8 are finite element approximations of the Michell structures solving those problems. For both problems there are available analytical solutions of Michell problem: see [100] for the cantilever problem and [144] for the bridge problem; the exact values of Z are quoted in Tables 4.1 and 4.2 in the rows corresponding to FibMD and $\mathrm{FibMD}_\pm$ settings. The shapes obtained herein by FE method resemble the exact *optimal layouts* given in those works very well; in each case $Z_h < Z$, which adheres to Proposition 4.44 (although Υ_0 is non-empty). With the FE mesh counting 2 250 000 elements the relative error $|Z - Z_h|/Z$ is around 0.2% for the cantilever problem and around 0.08% for the bridge problem.

Since the '60s, thanks to [75], another numerical method has been successfully used to approximate Michell structures. To simplify its presentation we shall assume that Ω is convex. The so called *ground structure* method invokes the underlying nature of Michell problem – from the universe of trusses that interconnect a finite, yet very large set of potential joints $\{A_i\}_{i=1}^m$ populating the design domain we choose the one with the least volume; it is crucial that positions of the joints are fixed. The name of the method refers to the "big" truss consisting of all $N = (m-1)\,m/2$ potential bars. A modified ground structure method will be the main numerical tool used in Chapters 5 and 6 therefore we shall now briefly review its original version.

With the generalized truss optimization problem (4.64) posed in Theorem 4.55 we may formulate the ground structure problem by simply restricting the support of measure $\mathscr{S}$ to a finite set. We shall adopt some notions and notations already introduced in Section 4.6 on the FE method. In particular, for the grid of potential joints or, as we shall henceforward call them, *nodes* we will take the set X^h of vertices of the FE mesh. It will be crucial that the load $F \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ is discretized to F_{X^h}, i.e. a measure supported on X^h. In 2D case to that aim we can use vectors $\mathbf{F}_1, \mathbf{F}_2 \in \mathbb{R}^m$ defined in (4.45):

$F_{\mathrm{X}^h} = \sum_{j=1}^{m} \big(\mathbf{F}_1(j)\,\mathrm{e}_1 + \mathbf{F}_2(j)\,\mathrm{e}_2\big)\,\delta_{v_j}$ where $\{v_1, \dots, v_m\} = \mathrm{X}^h \setminus \Upsilon_0$. Upon recalling Remark 4.56 the ground structure approximation of Michell problem may be given:

$$Z_{\mathrm{X}^h} = \min\Big\{ \int_{\mathrm{X}^h \times \mathrm{X}^h \setminus \Delta} |y - x| \Big(\frac{1}{\bar{s}_+} \mathscr{S}_+ + \frac{1}{\bar{s}_-} \mathscr{S}_- \Big)(dxdy) \ : \qquad (\mathrm{MP}_{\mathrm{X}^h})$$
$$\mathscr{S} \in \mathcal{M}(\mathrm{X}^h \times \mathrm{X}^h \setminus \Delta; \mathbb{R}),\ \mathcal{B}^* \mathscr{S} = F_{\mathrm{X}^h} \ \text{ in } \mathcal{M}^b(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d) \Big\}$$

By the lower index X^h we will distinguish the ground structure formulation from the FE formulation.

Proposition 4.57. *Let us assume convexity of Ω. The ground structure problem* $(\mathrm{MP}_{\mathrm{X}^h})$ *attains a solution provided that the set of admissible trusses $\mathscr{S}$ is non-empty. Moreover, if* $\operatorname{spt} F \subset \mathrm{X}^h$ *then it always holds that*

$$Z \le Z_{\mathrm{X}^h} \tag{4.67}$$

where $Z = \min \mathrm{MP} = \sup \mathrm{MP}^*$.

Remark 4.58. Considering the reviewing spirit of this section, to be brief we shall display the proof only for the case $\bar{s}_+ = \bar{s}_- = 1$ and $\Upsilon_0 = \varnothing$ when we can exploit Theorem 4.55.

Proof (for the case $\bar{s}_+ = \bar{s}_- = 1$ and $\Upsilon_0 = \varnothing$). Since X^h is finite, for any pair $(x, y) \in \mathrm{X}^h \times \mathrm{X}^h \setminus \Delta$ the distance $|x - y|$ is uniformly bounded from below by a positive constant hence the coerciveness of the minimized functional. The affine subspace of $\mathscr{S}$ satisfying $\mathcal{B}^* \mathscr{S} = F_{\mathrm{X}^h}$ is weakly-* closed and the existence result follows.

Once $\operatorname{spt} F \subset \mathrm{X}^h$ it is easy to check that $F_{\mathrm{X}^h} = F$ in $\mathcal{M}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$ (the load does not require discretization). Therefore the set of admissible trusses $\mathscr{S}$ in $(\mathrm{MP}_{\mathrm{X}^h})$ is a subset of the one in problem (4.64) hence inequality (4.67) follows due to the chain of equalities in Theorem 4.55. □

From the above result we see that in the case of a discrete load F the two methods: the finite element method formulated in Section 4.6.2.3 and the well known ground structure method furnish, respectively, lower and upper bounds for the exact volume of a Michell structure, namely $Z_h \le Z \le Z_{\mathrm{X}^h}$.

For the 2D case we shall now rewrite the ground structure problem $(\mathrm{MP}_{\mathrm{X}^h})$ to its algebraic form that is tailored for computer implementation. Let us recall that m stands for the number of nodes in $\mathrm{X}^h \setminus \Upsilon_0$ only. The number of bars in the ground structure equals $N = \big(\operatorname{card}(\mathrm{X}^h) - 1\big) \operatorname{card}(\mathrm{X}^h)/2$. We introduce functions $\mathrm{x}, \mathrm{y} : \{1, \dots, N\} \to \mathbb{R}^2$ such that the map $[\mathrm{x}(\,\cdot\,), \mathrm{y}(\,\cdot\,)] : \{1, \dots, N\} \to \big\{[x, y] \,:\, (x, y) \in \mathrm{X}^h \times \mathrm{X}^h \setminus \Delta\big\}$ is a bijection, i.e. $\mathrm{x}(k)$ and $\mathrm{y}(k)$ are the starting and the ending point of the k-th bar in the ground structure. We can discretize the geometric operator $\mathcal{B}$ given in (4.59) arriving at geometric matrices $\mathbf{B}_1, \mathbf{B}_2 \in \mathbb{R}^{N \times m}$ that are uniquely defined by relation

$$\big(\mathbf{B}_1 \mathbf{u}_1 + \mathbf{B}_2 \mathbf{u}_2\big)(k) = \Big(\mathcal{B}\big(\iota^h(\mathbf{u}_1, \mathbf{u}_2)\big)\Big)\Big(\mathrm{x}(k), \mathrm{y}(k)\Big), \tag{4.68}$$

where ι^h is the interpolation operator defined in Section 4.6 (any other interpolation operator will do as well since only the values of $\iota^h(\mathbf{u}_1, \mathbf{u}_2)$ on the set X^h matter). Next, vector $\mathbf{L} \in \mathbb{R}^N$ stores the information on the lengths of the bars, i.e.

$$\mathbf{L}(k) := \big|\mathrm{y}(k) - \mathrm{x}(k)\big|.$$

Let now the vectors $\mathbf{s}_+, \mathbf{s}_- \in \mathbb{R}^N_+$ represent the axial forces in the bars, i.e. to those vectors we relate the truss $\mathscr{S} = \sum_{k=1}^N \big(\mathbf{s}_+(k) - \mathbf{s}_-(k)\big)\, \delta_{(\mathrm{x}(k),\mathrm{y}(k))}$. By plugging this form of $\mathscr{S}$ into the ground structure problem ($\mathrm{MP}_{\mathrm{X}^h}$) we arrive at its algebraic form (we skip the details of the passage since it will be repeated in a similar setting in Section 5.5.4):

$$Z_{\mathrm{X}^h} = \min_{\mathbf{s}_+, \mathbf{s}_- \in \mathbb{R}^N_+} \left\{ \frac{1}{\bar{s}_+} \langle \mathbf{L}, \mathbf{s}_+ \rangle + \frac{1}{\bar{s}_-} \langle \mathbf{L}, \mathbf{s}_- \rangle \ : \ \begin{array}{l} \mathbf{B}_1^\top(\mathbf{s}_+ - \mathbf{s}_-) = \mathbf{F}_1 \\ \mathbf{B}_2^\top(\mathbf{s}_+ - \mathbf{s}_-) = \mathbf{F}_2 \end{array} \right\} \qquad (\mathbf{MP}_{\mathrm{X}^h})$$

We have thus recovered the ground structure formulation originally proposed in [75], for more contemporary texts see e.g. [157, 172, 31, 33]. It is a finite dimensional linear programming problem in its canonical form while its dual reads

$$Z_{\mathrm{X}^h} = \sup_{\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^m} \left\{ \langle \mathbf{u}_1, \mathbf{F}_1 \rangle + \langle \mathbf{u}_2, \mathbf{F}_2 \rangle \ : \ -\tfrac{1}{\bar{s}_-}\, \mathbf{L} \le \mathbf{B}_1\mathbf{u}_1 + \mathbf{B}_2\mathbf{u}_2 \le \tfrac{1}{\bar{s}_+}\, \mathbf{L} \right\} \qquad (\mathbf{MP}^*_{\mathrm{X}^h})$$

Connection of the problem above to the dual Michell problem (MP^*) can be easily found if characterization of $\overline{\mathcal{U}}_1$ from Theorem 4.55 is acknowledged (cf. Remark 4.56 for its generalization to $\bar{s}_+ \neq \bar{s}_-$). If $\operatorname{spt} F \subset \mathrm{X}^h$ through relation $u = \iota^h(\mathbf{u}_1, \mathbf{u}_2)$ we can see that the set of admissible displacement functions in problem ($\mathbf{MP}^*_{\mathrm{X}^h}$) is effectively larger than $\overline{\mathcal{U}}_1$: the two-point condition is checked on the set $\mathrm{X}^h \times \mathrm{X}^h \setminus \Delta$ only. We have thus obtained a dual argument for the inequality $Z \le Z_{\mathrm{X}^h}$.

Computational efficiency of the ground structure approach was greatly enhanced when in [91] an adaptive *member adding* algorithm was put forward and further developed in [158]. The idea consists in solving the problem iteratively for substructures of the full ground structure; in subsequent iterations new bars are added based on evaluation of the constraint in the dual problem ($\mathbf{MP}^*_{\mathrm{X}^h}$). The adaptive approach unlocks problems with N of order of magnitude equal to 10^9 which roughly corresponds to 201×201 nodal grid X^h. More details on the member adding technique will be explained in Section 5.5.5 where it will be originally adapted to the optimal membrane problem.

The ground structure approximations of Michell structures were computed for the cantilever and the bridge problem; note that in each case the load F is discrete. To that aim we used a program implemented in Mathematica that was developed in [158] and kindly shared by the author Tomasz Sokół himself. A ground structure interconnecting a 201×101 nodal grid X^h and counting $N = 206\,055\,150$ bars was used. The values Z_{X^h} are given in Tables 4.1 and 4.2 next to the exact values Z for the corresponding FibMD and $\mathrm{FibMD}_\pm$ settings of the (FMD) problem. We can see that $Z < Z_{\mathrm{X}^h}$ in each case while the relative error $|Z_{\mathrm{X}^h} - Z|/Z$ oscillates around 0.04% which compares favourably to FEM results (0.2% error). The results in the form of a truss are plotted for the cantilever problem only, cf. Fig. 4.9(a) for $\bar{s}_+ = \bar{s}_- = 1$ and Fig. 4.9(b) for $\bar{s}_+ = 1,\ \bar{s}_- = 1/3$; red and blue colours stand for bars in tension and compression respectively. To ensure better visual

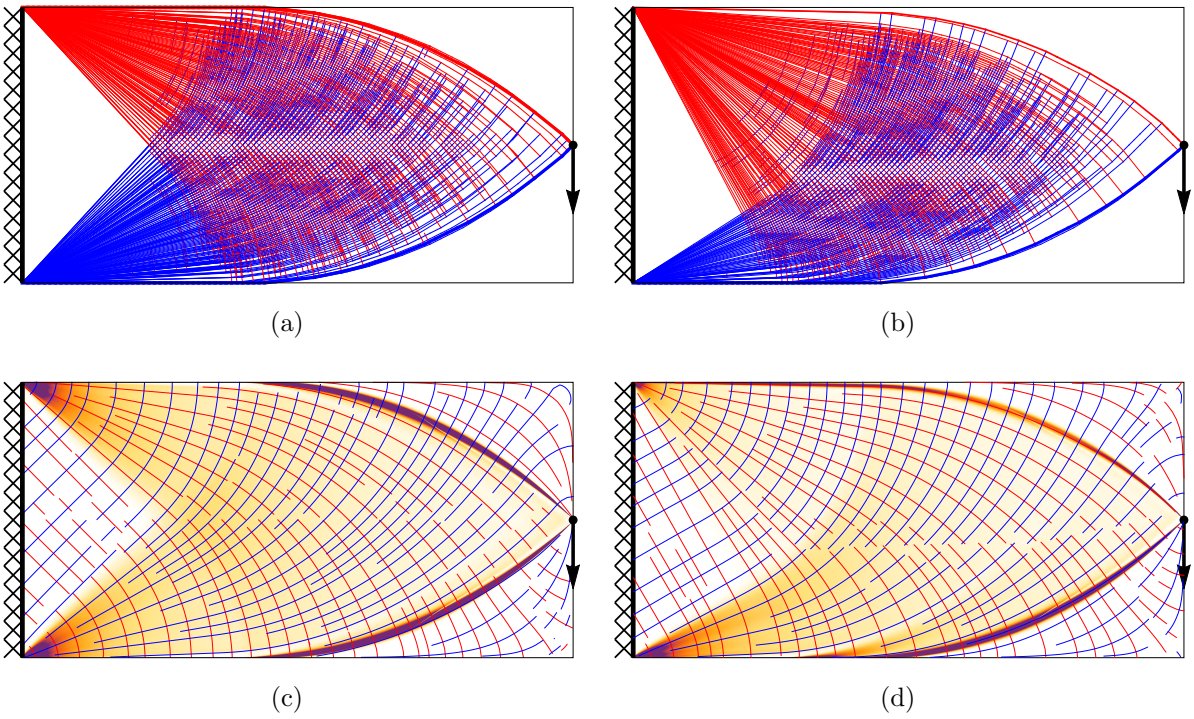


Figure 4.9: Trusses solving the ground structure approximation of Michell structure for the cantilever problem: (a) equal permissible stresses $\bar{s}_+ = \bar{s}_- = 1$; (b) non-equal permissible stresses $\bar{s}_+ = 1,\ \bar{s}_- = 1/3$. The bars in tension and compression are coloured red and blue respectively. Figures (c), (d) show the respective FE solutions along with trajectories of principal strains and stresses. The results (a), (b) were generated via computer program authored and kindly shared by Tomasz Sokół.

effects the pictures were produces for a more coarse ground structure generated by 101×51 nodal grid X^h. In Fig. 4.9(c),(d) the respective FEM solutions are presented; in order to make the comparison more crisp the trajectories of principles strains $e(u_h) = e\big(\iota^h(\mathbf{u}_1, \mathbf{u}_2)\big)$ were additionally plotted using an automatic procedure in Mathematica (they coincide with trajectories of principal stresses σ_h at points where material does not vanish). The visual correspondence between results computed via the two methods is clear.

In terms of the objective value Z the ground structure method seems to be superior to the FE method herein proposed, although a more comprehensive numerical study should be carried out to make such call. In a natural way the truss approximants can better bring out the 1D elements of Michell structures, i.e. the curves C satisfying $\mathcal{H}^1 \llcorner C \ll c_{\mathrm{M}}(\varsigma)$, which for the cantilever problem envelop the shape of the body. In contrast, the FE method seems to be more suited for mapping the part of ς that is absolutely continuous with respect to Lebesgue measure. Another advantage of the ground structure method is related to the equilibrium equations: for a discrete load F the resulting truss $\mathscr{S} = \sum_{k=1}^{N} \big(\mathbf{s}_+(k) - \mathbf{s}_-(k)\big)\, \delta_{(\mathrm{x}(k),\mathrm{y}(k))}$ induces the absolute stress $\varsigma_{\mathscr{S}} = \sum_{k=1}^{N} \big(\mathbf{s}_+(k) - \mathbf{s}_-(k)\big) \varsigma^{\mathrm{x}(k),\mathrm{y}(k)}$ that satisfies the exact equilibrium equation $-\mathrm{Div}\, \varsigma_{\mathscr{S}} = F$, cf. Proposition 4.54. This is not the case for the FE method were the resulting stress σ_h satisfy the approximate equilibrium equations only, i.e. $\sigma_h \in \Sigma^h(\Omega, F)$. With the FEM, on

the other hand, we obtain displacement field $u \in \iota^h(\mathbf{u}_1, \mathbf{u}_2)$ that rigorously satisfies $u \in \overline{\mathcal{U}}_1$ although it is not as relevant from the practical standpoint. It is also worth mentioning that the ground structure method creates a potential for 3D printing, especially if combined with geometric optimization algorithms, cf. [103], [104].

4.8. Monge-Kantorovich approach to optimal design of heat conductor

One of the key results of the paper [35] was the link between the Mass Optimization Problem (MOP) for a stationary heat conduction (i.e. when u is scalar and $\sigma = q$ is vectorial) and the *Optimal Transport Problem* of Monge and Kantorovich, cf. [166], [108]. In this section we show how to simply transfer this idea to the Free Material Design problem proposed herein.

4.8.1. The case $\Upsilon_0 = \varnothing$ – the classical Optimal Transport Problem

We consider the Free Material Design of heat conductor in accordance with Section 4.5.3 with additional assumptions on the data:

$$\Omega \text{ is convex}, \qquad \Upsilon_0 = \varnothing, \qquad \int_{\overline{\Omega}} df = 0. \tag{4.69}$$

In respect of Remark 4.43 it is general enough to assume the AMD setting of (FMD) in which $j(H, \xi) = \frac{1}{2}\langle H\xi, \xi\rangle$, the admissible set $\mathcal{K} = \mathcal{S}^{d\times d}_+$ contains all positive semi-definite conductivity tensors and $c(H) = \operatorname{Tr} H$. Such (FMD) problem has been established to reduce to the pair of problems $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$ for the functions $\rho = \rho^0 = |\cdot|$ being Euclidean norms on $\mathbb{R}^d$.

According to Proposition 4.40 the set $\overline{\mathcal{U}}_1$ consists of Lipschitz continuous functions for which $\rho(\nabla u) = |\nabla u| \leq 1$ a.e. in Ω. Since Ω is convex this condition is equivalent to bounding the Lipschitz constant of the function: $\operatorname{lip}(u) \leq 1$, cf. [8, Proposition 2.13]. Problem $(\overline{\mathcal{P}})$ may thus be rewritten:

$$Z = \max\left\{ \int u\, df \ : \ u \in C(\overline{\Omega}), \ \ u(y) - u(x) \leq |y - x| \quad \forall\, (x, y) \in \overline{\Omega} \times \overline{\Omega} \right\}. \tag{4.70}$$

Let us now jump to pose the Monge-Kantorovich Optimal Transport Problem (OTP). Let $\nu_1, \nu_2 \in \mathcal{M}_+(\overline{\Omega})$ be two finite positive measures of equal mass, i.e. $\nu_1(\overline{\Omega}) = \nu_2(\overline{\Omega})$ (two probability measures on $\overline{\Omega}$ are the standard choice). The two measures ν_1 and ν_2 shall be termed *source measure* and *target measure* respectively. We define the set of feasible *transportation plans*:

$$\Gamma(\nu_1, \nu_2) := \left\{ \gamma \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega}) \ : \ \begin{array}{l} \gamma(B \times \overline{\Omega}) = \nu_1(B) \\ \gamma(\overline{\Omega} \times B) = \nu_2(B) \end{array} \text{ for any Borel set } B \subset \overline{\Omega} \right\}.$$

We say that condition $\gamma \in \Gamma(\nu_1, \nu_2)$ fixes the left and the right marginal of γ as ν_1 and ν_2 respectively, which can be also written by means of push-forward measures (see Section 2.2) with respect to the projection mappings: $\pi_1 \# \gamma = \nu_1$ and $\pi_2 \# \gamma = \nu_2$ where $\pi_1(x, y) = x$,

$\pi_2(x,y)=y$. From [166] we learn that a measure $\gamma\in\mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$ is an element of $\Gamma(\nu_1,\nu_2)$ if and only if

$$\int_{\overline{\Omega}\times\overline{\Omega}}\big(\varphi(x)+\psi(y)\big)\,\gamma(dxdy)=\int_{\overline{\Omega}}\varphi\,d\nu_1+\int_{\overline{\Omega}}\psi\,d\nu_2\qquad\forall\,\varphi,\psi\in C(\overline{\Omega}).\tag{4.71}$$

We observe that $\gamma(\overline{\Omega}\times\overline{\Omega})=\nu_1(\overline{\Omega})=\nu_2(\overline{\Omega})$ for any γ in $\Gamma(\nu_1,\nu_2)$ implying weak-* pre-compactness of the latter set. Then, weak-* compactness of $\Gamma(\nu_1,\nu_2)$ follows easily by the characterization above. The set is also non-empty since $\gamma=\nu_1\otimes\nu_2$ is its element.

By the Optimal Transport Problem of Monge and Kantorovich (for the distance cost) one understands the problem of finding a transportation plan γ between ν_1 and ν_2 that yields the minimal transport cost:

$$W(\nu_1,\nu_2):=\min\bigg\{\int_{\overline{\Omega}\times\overline{\Omega}}|y-x|\,\gamma(dxdy)\;:\;\gamma\in\Gamma(\nu_1,\nu_2)\bigg\}\tag{OTP}$$

The objective functional is linear continuous on $\mathcal{M}(\overline{\Omega}\times\overline{\Omega})$ therefore existence of solution of (OTP) follows by compactness and non-emptiness of $\Gamma(\nu_1,\nu_2)$. For probability measures $\nu_1,\nu_2\in\mathrm{P}(\overline{\Omega})$ the value $W(\nu_1,\nu_2)$ is known as the *Wasserstein distance* between ν_1 and ν_2. The dual formulation of the Monge-Kantorovich problem reads

$$W(\nu_1,\nu_2)=\sup\bigg\{\int_{\overline{\Omega}}\varphi\,d\nu_1+\int_{\overline{\Omega}}\psi\,d\nu_2\;:\qquad\qquad\tag{OTP*}$$
$$\varphi,\psi\in C(\overline{\Omega}),\;\varphi(x)+\psi(y)\le|y-x|\;\;\forall\,(x,y)\in\overline{\Omega}\times\overline{\Omega}\bigg\}$$

cf. [166, Theorem 1.3]. Connection between (OTP) and $(\overline{\mathcal{P}})$ may readily be given:

Theorem 4.59 (**Kantorovich-Rubinstein**). *Let there be given source and target measures $\nu_1,\nu_2\in\mathcal{M}_+(\overline{\Omega})$ such that $f=\nu_1-\nu_2$ ($\nu_1=f_+$, $\nu_2=f_-$ is a possible choice). Then, under assumptions* (4.69) *there hold equalities*

$$Z=W(\nu_1,\nu_2)=W(f_+,f_-).$$

Moreover, for any solution u of $(\overline{\mathcal{P}})$ functions $\varphi=u$, $\psi=-u$ solve (OTP*).

Proof. See [166, Theorem 1.14]. □

Remark 4.60. The Kantorovich-Rubenstein theorem remains true if the Euclidean distance $|y-x|$ is replaced by a cost $c(x,y)$ provided that c is any lower semi-continuous pseudo-metric (a non-negative, symmetric and sub-additive function) on $\overline{\Omega}$.

Remark 4.61. From the Kantorovich-Rubenstein result we can see that $W(\nu_1,\nu_2)$ depends on the difference $\nu_1-\nu_2$ only, therefore $W(\nu_1,\nu_2)=W(\nu_1+\nu_0,\nu_2+\nu_0)$ for any $\nu_0\in\mathcal{M}_+(\overline{\Omega})$.

To find a dual connection between the (FMD) problem and the Monge-Kantorovich problem, i.e. between problems $(\mathcal{P}^*)$ and (OTP), to each transportation plan we will associate the (negative) heat flux $\vartheta_\gamma\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$. We start by defining a unit heat flux concentrated on a segment: for any pair $(x,y)\in\overline{\Omega}\times\overline{\Omega}$

$$\vartheta^{x,y}:=\tau^{x,y}\,\mathcal{H}^1\llcorner[x,y]$$

where $\tau^{x,y} = \frac{y-x}{|y-x|}$ when x, y are distinct and $\tau^{x,y} = 0$ otherwise. For any $\gamma \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ the map $(x, y) \mapsto \vartheta^{x,y}$ is γ-measurable and (thanks to boundedness of $\overline{\Omega}$) γ-essentially bounded in the sense of Definition 2.26. Then, we may define a Radon measure ϑ_γ through the integration formula:

$$\langle \epsilon, \vartheta_\gamma \rangle_{(C,\mathcal{M})} := -\int_{\overline{\Omega}\times\overline{\Omega}} \left(\int \langle \epsilon, \vartheta^{x,y} \rangle \right) \gamma(dxdy) \qquad \forall\, \epsilon \in C(\overline{\Omega}; \mathbb{R}^d). \tag{4.72}$$

In the nomenclature of Monge-Kantorovich problem measures $\vartheta^{x,y}$ (or rather the segments $[x, y]$ themselves) are called *transport rays*. For the intuition behind the minus sign in the definition (4.72), see Remark 4.41.

Proposition 4.62. *For a convex domain Ω let $\nu_1, \nu_2 \in \mathcal{M}_+(\overline{\Omega})$ be any measures of equal mass. Then, for any transportation plan $\gamma \in \Gamma(\nu_1, \nu_2)$ the Radon measure ϑ_γ is supported on $\overline{\Omega}$ while equation $-\operatorname{div} \vartheta_\gamma = \nu_1 - \nu_2$ is satisfied in $\mathcal{D}'(\mathbb{R}^d)$.*

Proof. The fact that $\operatorname{spt} \vartheta_\gamma \subset \overline{\Omega}$ follows from boundedness and convexity of Ω. For any $\phi \in \mathcal{D}(\mathbb{R}^d)$ by performing integration by parts as in the proof of Proposition 4.54 we end up with

$$\langle \phi, -\operatorname{div} \vartheta_\gamma \rangle_{(\mathcal{D},\mathcal{D}')} = -\int_{\overline{\Omega}\times\overline{\Omega}} \big(\phi(y) - \phi(x)\big)\, \gamma(dxdy) = \int_{\overline{\Omega}} \phi\, d\nu_1 - \int_{\overline{\Omega}} \phi\, d\nu_2$$

where in the last equality we used the fact that $\gamma \in \Gamma(\nu_1, \nu_2)$, see (4.71). □

We are ready to recover the well-known result on Monge-Kantorovich problem, see e.g. [35, 76, 166]:

Proposition 4.63. *For a convex bounded design domain $\Omega \subset \mathbb{R}^d$ let $\gamma \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ denote any solution of* (OTP) *for any $\nu_1, \nu_2 \in \mathcal{M}_+(\overline{\Omega})$ such that $\nu_1 - \nu_2 = f$.*

Then, the measure $\vartheta_\gamma \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ solves problem $(\mathcal{P}^*)$.

Proof. Admissibility of ϑ_γ in $(\mathcal{P}^*)$ follows by virtue of Proposition 4.62. For any $\epsilon \in C(\overline{\Omega}; \mathbb{R}^d)$ satisfying $\rho(\epsilon) = |\epsilon| \le 1$ in $\overline{\Omega}$

$$\int \langle \epsilon, \vartheta_\gamma \rangle = \int \left(\int_{[x,y]} \langle -\epsilon, \vartheta^{x,y} \rangle\, d\mathcal{H}^1 \right) \gamma(dxdy) \le \int |y - x|\, \gamma(dxdy) = W(\nu_1, \nu_2)$$

where the last equality exploits optimality of γ. By Corollary 2.24 we have $\rho^0(\vartheta_\gamma) = \sup_{\epsilon \in C(\overline{\Omega};\mathbb{R}^d)} \big\{ \int \langle \epsilon, \vartheta_\gamma \rangle \,:\, \rho(\epsilon) \le 1 \ \text{ in } \overline{\Omega} \big\}$ and by Theorem 4.59 we deduce that $\int \rho^0(\vartheta) \le Z$ establishing optimality of ϑ. □

The construction of solution ϑ to $(\mathcal{P}^*)$ based on solution of (OTP) paves a way towards solution λ of the (FMD) problem itself: by virtue of Corollary 4.7 and characterization of the set $\underline{\mathcal{K}}_1(q)$ optimality of the field $\lambda = \frac{C_0}{Z}\big(\frac{d\vartheta}{d|\vartheta|} \otimes \frac{d\vartheta}{d|\vartheta|}\big)|\vartheta|$ follows, cf. Section 4.5.3. Nonetheless, this formula drops a useful feature of decomposability of the tensor-valued measure λ to family of rank-one 1D measures $\tau^{x,y} \otimes \tau^{x,y}\, \mathcal{H}^1 \llcorner [x, y]$. Putting it differently, one may *a priori* expect superposition of measures $\vartheta^{x,y}$ on a set of non-zero measure $\mu = \operatorname{Tr} \lambda$ which could cause $\frac{d\vartheta}{d|\vartheta|}$ to rotate with respect to the segments. By virtue of a well known theorem on Monge-Kantorovich problem, however, "transport rays do not cross or overlap" (see [81]) thus excluding possibility of such superposition. Notwithstanding

this, the decomposability of λ can be deduced more directly, without employing this deep result:

Theorem 4.64. *For transportation plan γ satisfying the prerequisite of Proposition 4.63 let us define the conductivity tensor field $\lambda_\gamma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ via the formula*

$$\langle \Psi, \lambda_\gamma \rangle_{(C,\mathcal{M})} := \frac{C_0}{Z} \int_{\overline{\Omega}\times\overline{\Omega}} \left(\int_{[x,y]} \langle \Psi(s), \tau^{x,y} \otimes \tau^{x,y} \rangle \, \mathcal{H}^1(ds) \right) \gamma(dxdy) \qquad \forall\, \Psi \in C(\overline{\Omega}; \mathcal{S}^{d\times d}). \tag{4.73}$$

Then λ_γ is a solution of the (FMD) *problem for stationary heat conduction with $\Upsilon_0 = \varnothing$.*

Proof. First we must verify whether λ_γ is feasible for (FMD). We see that $\operatorname{Tr} \lambda_\gamma = \langle \mathrm{I}, \lambda_\gamma \rangle$ where I is the identity $d \times d$ matrix. If I is treated as a constant function in $C(\overline{\Omega}; \mathcal{S}^{d\times d})$ then

$$\begin{aligned} \int \operatorname{Tr} \lambda_\gamma = \langle \mathrm{I}, \lambda_\gamma \rangle_{(C,\mathcal{M})} &= \frac{C_0}{Z} \int_{\overline{\Omega}\times\overline{\Omega}} \left(\int_{[x,y]} \langle \mathrm{I}, \tau^{x,y} \otimes \tau^{x,y} \rangle \, d\mathcal{H}^1 \right) \gamma(dxdy) \\ &= \frac{C_0}{Z} \int_{\overline{\Omega}\times\overline{\Omega}} |x-y| \, \gamma(dxdy) = C_0, \end{aligned}$$

where we used the fact that γ solves (OTP) and $W(f_+, f_-) = Z$; feasibility of λ_γ in (FMD) is validated.

Let $\vartheta_\gamma \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ be the vector measure given by formula (4.72); owing to Proposition 4.62 it satisfies $-\operatorname{div} \vartheta_\gamma = f$ or equivalently $\int u \, df = \int \langle \nabla u, \vartheta_\gamma \rangle$ for any $u \in \mathcal{D}(\mathbb{R}^d)$. Since $j(H, \xi) = \frac{1}{2}\langle H\xi, \xi \rangle$ compliance $\mathcal{C}(\lambda_\gamma)$ can be thus rewritten and estimated as follows

$$\begin{aligned} &\mathcal{C}(\lambda_\gamma) = \sup \left\{ \int \langle \nabla u, \vartheta_\gamma \rangle - \frac{1}{2} \int \langle \lambda_\gamma, \nabla u \otimes \nabla u \rangle \ : \ u \in \mathcal{D}(\mathbb{R}^d) \right\} \\ &\leq \sup \left\{ \int \langle \epsilon, \vartheta_\gamma \rangle - \frac{1}{2} \int \langle \lambda_\gamma, \epsilon \otimes \epsilon \rangle \ : \ \epsilon \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^d) \right\} \\ &= \sup \left\{ \int_{\overline{\Omega}\times\overline{\Omega}} \left(\int_{[x,y]} \left(\langle \epsilon(s), \tau^{x,y} \rangle - \frac{C_0}{2Z} \Big(\langle \epsilon(s), \tau^{x,y} \rangle \Big)^2 \right) \mathcal{H}^1(ds) \right) \gamma(dxdy) : \epsilon \in \mathcal{D}(\mathbb{R}^d; \mathbb{R}^d) \right\} \\ &\leq \int_{\overline{\Omega}\times\overline{\Omega}} \left(\int_{[x,y]} \frac{Z}{2C_0} \, \mathcal{H}^1(dz) \right) \gamma(dxdy) = \frac{Z}{2C_0} \int_{\overline{\Omega}\times\overline{\Omega}} |x-y| \, \gamma(dxdy) = \frac{Z^2}{2C_0}, \end{aligned}$$

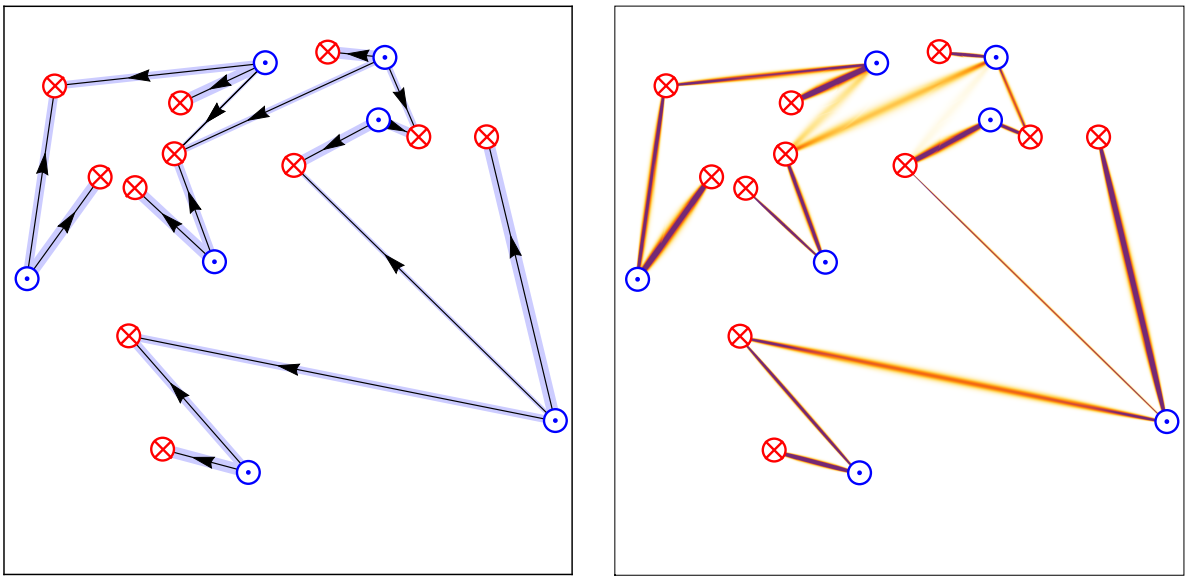
where to obtain the last inequality we used the fact that $\sup_{t\in\mathbb{R}} \{ t - \frac{C_0}{2Z} t^2 \} = \frac{Z}{2C_0}$. The last term in the chain above: $\frac{Z^2}{2C_0}$ is precisely the value of minimum compliance $\mathcal{C}_{\min}$ (cf. Lemma 4.2 for $p = 2$) proving that λ_γ solves (FMD). □

Example 4.65 (**Barrett-Prigozhin coupling problem**)**.** Let us consider the Free Material Design of conductor that within a square domain $\Omega = (0,1)^2$ ought to transfer heat from a 7-point source $f_+ = \sum_{i=1}^{7} \frac{1}{7} \delta_{x_i}$ to an 11-point sink $f_- = \sum_{j=1}^{11} \frac{1}{11} \delta_{y_j}$, namely $f = f_+ - f_-$; for location of points x_i, y_j see Fig. 4.10. The corresponding Monge-Kantorovich problem (OTP) for the probability source and target measures $\nu_1 = f_+$ and, respectively, $\nu_2 = f_-$ was originally considered in [11, Example 3] (exact positions of x_i, y_j may slightly differ).

For this discrete (OTP) the constraints on marginals automatically restrict the support of feasible transportation plan to a finite set: $\operatorname{spt}\gamma \subset \{x_1,\dots,x_7\}\times\{y_1,\dots,y_{11}\}$. More accurately $\gamma\in\Gamma(f_+,f_-)$ if and only if

$$\gamma=\sum_{i=1}^{7}\sum_{j=1}^{11}\gamma_{ij}\,\delta_{(x_i,y_j)},\qquad \gamma_{ij}\geq 0,\qquad \sum_{j=1}^{11}\gamma_{ij}=\frac{1}{7}\quad\forall i,\qquad \sum_{i=1}^{7}\gamma_{ij}=\frac{1}{11}\quad\forall j$$

while $\int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta}|y-x|\gamma(dxdy)=\sum_{i=1}^{7}\sum_{j=1}^{11}|y_j-x_i|\,\gamma_{ij}$, rendering (OTP) a finite dimensional linear programming problem that in [11] is referred to as *a coupling problem*. The exact solution is showed in Fig. 4.10(a); the widths of the segments are proportional to γ_{ij} while the arrows indicate direction of the transport and, in fact, of the heat flux $-\vartheta_\gamma$. The exact minimum transport cost equals $W(f_+,f_-)\approx 0.25585$ which is also the value of $Z=\max(\overline{\mathcal{P}})=\min(\mathcal{P}^*)$. The conductivity tensor field λ_γ in accordance with Theorem 4.64 yields a solution of the (FMD) problem.



(a) (b)

Figure 4.10: The coupling problem formulated in [11]; $f=\sum_{i=1}^{7}\frac{1}{7}\,\delta_{x_i}-\sum_{j=1}^{11}\frac{1}{11}\,\delta_{y_j}$, the source points x_i are denoted by circled points and the target points y_j by circled crosses. (a) The exact solution of the Monge-Kantorovich problem via linear programming, $W(f_+,f_-)=Z\approx 0.25585$; (b) the FE approximation for a 4 500 000-element mesh – the heat-map of the heat flux density $|q_h|$ solving $(\mathcal{P}_h^*)$ (and of $\operatorname{Tr}\mathscr{H}_h$ optimal for (FMD_h)) with $Z_h=0.25573$.

The work [11] puts forward a mixed formulation of a finite element method for Monge-Kantorovich problem or, virtually, for the pair of problems $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$. The FE analysis is directly carried out on a regularization of the pair where instead of $\rho^0(q)=|q|$ the integrand $\frac{1}{r}|q|^r$ for r slightly bigger than one is considered. The authors use their method to find an approximate solution of the coupling problem; with their mesh adaptation technique they end up with a 96 466 element mesh and an approximate transport cost $W^h(f_+,f_-)$ that

deviates from the exact value by 0.34%. Fig. 4.10(b) shows the FE approximation obtained by the conic programming formulation proposed in this thesis (the elasticity framework in Section 4.6 is easily translated to conductivity). Computation was performed for a 4 500 000 mesh, which has taken 60 min. 4 s and produced $W^h(f_+, f_-) := Z^h = 0.25573$ yielding a 0.05% relative error. The visual comparison with the result in [11] seems to be in favour of the one herein presented – the transport rays appear to be more dispersed in [11]. Those differences, however, are rather due to the difference in mesh resolutions than in the methods themselves – the one proposed here is much less sophisticated, in particular the flux density q_h is guaranteed to be of higher regularity in [11]. △

4.8.2. The case $\Upsilon_0 = \partial\Omega$ and optimal transport of mass to the boundary

For the case when the Dirichlet zone Υ_0 is non-empty the relationship between optimal design of conductor and Monge-Kantorovich problem becomes less obvious. For the Mass Optimization Problem the work [35] shows that this issue is related to finding in (OTP) optimal hidden source and target measures ν_1^0 and ν_2^0 supported on Υ_0. Once again we shall bring those ideas to the Free Material Design of conductor. We will limit ourselves to the scenario when $\Upsilon_0 = \partial\Omega$, which is directly related to the optimal membrane problem put forward in Chapter 5, cf. the link studied in Section 5.6. It will be not restrictive to assume that $|f|(\partial\Omega) = 0$ and similarly $\nu_1(\partial\Omega) = \nu_2(\partial\Omega) = 0$.

Below we agree that (FMD) and $Z = \max(\overline{\mathcal{P}}) = \min(\mathcal{P}^*)$ refer to the heat conductor design problem considered in the previous subsection only with $\Upsilon_0 = \partial\Omega$; in particular combining Proposition 4.40 with [8, Proposition 2.13] furnishes the characterization

$$\overline{\mathcal{U}}_1 = \Big\{ u \in C_0(\Omega) \,:\, u(y) - u(x) \leq |y - x| \quad \forall\, (x, y) \in \overline{\Omega} \times \overline{\Omega} \Big\},$$

note that $u \in C_0(\Omega)$ is equivalent to enforcing $u = 0$ on $\partial\Omega$ for $u \in C(\overline{\Omega})$.

By the *Optimal Transport Problem with a fee free zone* $\partial\Omega$ we will understand the problem of designing the transportation plan γ along with source and target measures ν_1^0 and, respectively, ν_2^0 that are supported on $\partial\Omega$:

$$\begin{aligned} W^{\partial\Omega}(\nu_1, \nu_2) &:= \inf_{\nu_1^0, \nu_2^0 \in \mathcal{M}_+(\partial\Omega)} W(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) \\ &= \inf_{\substack{\nu_1^0, \nu_2^0 \in \mathcal{M}_+(\partial\Omega) \\ \gamma \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})}} \left\{ \int_{\overline{\Omega} \times \overline{\Omega}} |x - y| \, \gamma(dxdy) \;:\; \gamma \in \Gamma(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) \right\}. \end{aligned} \tag{OTP$^{\partial\Omega}$}$$

Note that the set of admissible triples ν_1^0, ν_2^0, γ is not *a priori* compact therefore existence of solution to (OTP$^{\partial\Omega}$) is not clear. Nevertheless, combining the Kantorovich-Rubinstein theorem with Ky Fan's result on switching inf and sup furnishes:

Lemma 4.66. *For $\Upsilon_0 = \partial\Omega$ and any $\nu_1, \nu_2 \in \mathcal{M}_+^b(\Omega)$ such that $\nu_1 - \nu_2 = f$ there holds an equality $W^{\partial\Omega}(\nu_1, \nu_2) = Z$, namely*

$$W^{\partial\Omega}(\nu_1, \nu_2) = \max \Big\{ \int u \, df \;:\; u \in C_0(\Omega), \quad u(y) - u(x) \leq |x - y| \quad \forall\, (x, y) \in \overline{\Omega} \times \overline{\Omega} \Big\}.$$

Moreover, ($\mathrm{OTP}^{\partial\Omega}$) *always admits solution* ν_1^0, ν_2^0, γ *and one can enforce that* $\nu_1^0(\partial\Omega) \leq \nu_2(\Omega)$ *and* $\nu_2^0(\partial\Omega) \leq \nu_1(\Omega)$.

The proof of the lemma is postponed to Section 5.6 where it shall be formulated for a more general setting, cf. Lemma 5.85. With the use of the lemma above we immediately obtain a variation of Theorem 4.64:

Theorem 4.67. *For a convex bounded design domain* $\Omega \subset \mathbb{R}^d$ *let* $(\nu_1^0, \nu_2^0, \gamma)$ *denote any solution of* ($\mathrm{OTP}^{\partial\Omega}$) *for any* $\nu_1, \nu_2 \in \mathcal{M}_+^b(\Omega)$ *such that* $\nu_1 - \nu_2 = f$.

Then, the conductivity tensor field $\lambda_\gamma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}_+^{d\times d})$ *defined via integration formula in Theorem 4.64 solves* (FMD) *problem for stationary heat conduction with* $\Upsilon_0 = \partial\Omega$.

Proof. By Proposition 4.62 we have $-\mathrm{div}\,\vartheta_\gamma = (\nu_1 + \nu_1^0) - (\nu_2 + \nu_2^0) = f + (\nu_1^0 - \nu_2^0)$ in $\mathcal{D}'(\mathbb{R}^d)$ and thus $-\mathrm{div}\,\vartheta_\gamma = f$ in $\mathcal{D}'(\mathbb{R}^d \setminus \partial\Omega)$ (and in fact in $\mathcal{D}'(\Omega)$). The rest of the argument boils down to repeating the proof of Theorem 4.64 while acknowledging Lemma 4.66. □

One can observe that the condition on equality of masses $\nu_1(\Omega) = \nu_2(\Omega)$ is not obligatory for well-posedness of ($\mathrm{OTP}^{\partial\Omega}$), contrarily to the classical (OTP). In particular one may consider $\nu_1 = f$ and $\nu_2 = 0$ once f is a positive bounded measure on Ω; further we will assume that $f \in \mathrm{P}(\Omega)$, i.e. that f is a probability. In this scenario ($\mathrm{OTP}^{\partial\Omega}$) becomes a problem of *optimal transport of mass* f *to the boundary*. Indeed, by exploiting the condition $\nu_1^0(\partial\Omega) \leq \nu_2(\Omega) = 0$ from Lemma 4.66 we obtain

$$\begin{aligned} W^{\partial\Omega}(f) := W^{\partial\Omega}(f,0) &= \min_{\nu_2^0 \in \mathrm{P}(\partial\Omega)} W(f, \nu_2^0) \\ &= \min_{\substack{\nu_2^0 \in \mathrm{P}(\partial\Omega) \\ \gamma \in \mathrm{P}(\overline{\Omega}\times\overline{\Omega})}} \left\{ \int_{\overline{\Omega}\times\overline{\Omega}} |x-y|\, \gamma(dxdy) \ : \ \gamma \in \Gamma(f, \nu_2^0) \right\} \\ &= \min_{\gamma \in \mathrm{P}(\overline{\Omega}\times\overline{\Omega})} \left\{ \int_{\overline{\Omega}\times\partial\Omega} |x-y|\, \gamma(dxdy) \ : \ \mathrm{spt}(\gamma) \subset \overline{\Omega}\times\partial\Omega, \ \ \pi_1\#\gamma = f \right\}, \end{aligned}$$

namely the second marginal $\pi_2\#\gamma$ is an arbitrary probability on $\partial\Omega$. We propose that the value $W^{\partial\Omega}(f)$ be termed *Wasserstein distance of mass* f *to the boundary*.

We shall now show that for $\nu_1 = f \in \mathrm{P}(\Omega)$ and $\nu_2 = 0$ the ($\mathrm{OTP}^{\partial\Omega}$) boils down to a geometric problem related to *sandpile modelling* and *eikonal equation*, see e.g. [50]. For each $x \in \overline{\Omega}$ let us introduce

$$d^{\partial\Omega}(x) := \inf_{y\in\partial\Omega} |y-x|, \qquad \Pi^{\partial\Omega}(x) := \left\{ y \in \partial\Omega \ : \ |y-x| = d^{\partial\Omega}(x) \right\} \tag{4.74}$$

It is straightforward to see that $d^{\partial\Omega}$ is non-negative and equals zero on $\partial\Omega$ and only there. Since $\partial\Omega$ is compact the minimum in definition of $d^{\partial\Omega}(x)$ is achieved hence the non-emptiness of $\Pi^{\partial\Omega}(x)$ for every x. Next we check that $d^{\partial\Omega}$ is Lipschitz continuous with $\mathrm{lip}(u) \leq 1$. Indeed, for $x_1, x_2 \in \overline{\Omega}$ let us choose $y_2 \in \partial\Omega$ such that $|y_2 - x_2| = d^{\partial\Omega}(x_2)$, then by the triangle inequality $d^{\partial\Omega}(x_1) - d^{\partial\Omega}(x_2) = d^{\partial\Omega}(x_1) - |y_2 - x_2| \leq |y_2 - x_1| - |y_2 - x_2| \leq |x_1 - x_2|$. We thus infer that $d^{\partial\Omega} \in \overline{\mathcal{U}}_1$. The multifunction $\Pi^{\partial\Omega}$ to each x maps the subset of those points on $\partial\Omega$ which are closest to x. We define the graph of $\Pi^{\partial\Omega}$:

$$\mathcal{G}^{\partial\Omega} := \left\{ (x,y) \in \overline{\Omega}\times\partial\Omega \ : \ y \in \Pi^{\partial\Omega}(x) \right\}.$$

This set can be rewritten as $\mathcal{G}^{\partial\Omega} = \big\{(x,y) \in \overline{\Omega}\times\partial\Omega \,:\, |x-y| = d^{\partial\Omega}(x)\big\}$ and by continuity of $d^{\partial\Omega}$ we can see that $\mathcal{G}^{\partial\Omega}$ is closed, which makes the probability space $\mathrm{P}(\mathcal{G}^{\partial\Omega})$ well defined. Moreover, closedness of $\mathcal{G}^{\partial\Omega}$ is equivalent to lower semi-continuity of the multifunction $\Pi^{\partial\Omega}$, see Theorem 2.31. After Theorems 2.36 and 2.37 we deduce that there always exists a Borel measurable selection $\pi_{\partial\Omega} : \Omega \to \partial\Omega$ of multifunction $\Pi^{\partial\Omega}$; every such function $\pi_{\partial\Omega}$ can be considered an orthogonal projection onto $\partial\Omega$. One can check that

$$\gamma = (\mathrm{id}\times\pi_{\partial\Omega})\#f \quad \in \quad \Big\{\gamma \in \mathrm{P}(\mathcal{G}^{\partial\Omega}) \,:\, \pi_1\#\gamma = f\Big\} \tag{4.75}$$

where $(\mathrm{id}\times\pi_{\partial\Omega}) : \overline{\Omega} \ni x \mapsto \big(x, \pi_{\partial\Omega}(x)\big) \in \mathcal{G}^{\partial\Omega}$. Non-emptiness of the set above justifies the following construction of optimal transport of mass to the boundary:

Proposition 4.68. *For a source $f \in \mathrm{P}(\Omega)$ let us take any pairing measure $\gamma \in \mathrm{P}(\mathcal{G}^{\partial\Omega})$ with the left margin $\pi_1\#\gamma$ equal to f.*

Then the triple γ, $\nu_1^0 = 0$, $\nu_2^0 = \pi_2\#\gamma$ solves $(\mathrm{OTP}^{\partial\Omega})$ for $\nu_1 = f$, $\nu_2 = 0$ while the Wasserstein distance of f to boundary $\partial\Omega$ equals

$$W^{\partial\Omega}(f) = \int_{\overline{\Omega}} d^{\partial\Omega}(x)\, f(dx)$$

and $u = d^{\partial\Omega}$ solves $(\overline{\mathcal{P}})$.

Proof. We start by proving that $u = d^{\partial\Omega}$ solves $(\overline{\mathcal{P}})$; we have already established that $u \in \overline{\mathcal{U}}_1$. We can see that for any $x_0 \in \overline{\Omega}$ there holds $\tilde{u}(x_0) \le u(x_0)$ for every $\tilde{u} \in \overline{\mathcal{U}}_1$. Indeed, let us choose $y_0 \in \partial\Omega$ such that $|y_0 - x_0| = d^{\partial\Omega}(x_0)$; then due to the two-point characterization of $\overline{\mathcal{U}}_1$ we have $\tilde{u}(x_0) \le |x_0 - y_0| + \tilde{u}(y_0) = d^{\partial\Omega}(x_0) = u(x_0)$ since $\tilde{u} = 0$ on $\partial\Omega$. The function u thus attains maximal possible value at all points of $\overline{\Omega}$ altogether therefore its optimality follows by positivity of f. By Lemma 4.66 we have $W^{\partial\Omega}(f) = W^{\partial\Omega}(f,0) = Z = \int u\,df = \int d^{\partial\Omega}(x)\, f(dx)$. Clearly $\mathrm{spt}(\gamma) \subset \mathcal{G}^{\partial\Omega} \subset \overline{\Omega}\times\partial\Omega$ while

$$\int_{\mathcal{G}^{\partial\Omega}} |x-y|\,\gamma(dxdy) = \int_{\mathcal{G}^{\partial\Omega}} d^{\partial\Omega}(x)\,\gamma(dxdy) = \int_{\overline{\Omega}} d^{\partial\Omega}(x)\,\gamma(dxdy) = \int_{\overline{\Omega}} d^{\partial\Omega}(x)\, f(dx)$$

where the last equality is due to $\pi_1\#\gamma = f$, cf. the characterization in (4.71). We have therefore established that $\int_{\mathcal{G}^{\partial\Omega}} |x-y|\,\gamma(dxdy) = W^{\partial\Omega}(f) = W^{\partial\Omega}(f,0)$ proving optimality of the triple γ, ν_1^0, ν_2^0 for $(\mathrm{OTP}^{\partial\Omega})$. □

Remark 4.69. Essentially Proposition 4.68 fully characterizes the set of solutions of $(\mathrm{OTP}^{\partial\Omega})$: one can easily show that γ cannot be optimal when $\mathrm{spt}(\gamma) \not\subseteq \mathcal{G}^{\partial\Omega}$.

In general the condition $\gamma \in \mathrm{P}(\mathcal{G}^{\partial\Omega})$ allows dispatching the source mass f to different parts of $\partial\Omega$. The plan $\gamma = (\mathrm{id}\times\pi_{\partial\Omega})\#f$, however, guarantees that the whole mass at each point x is transported to the one point $\pi_{\partial\Omega}(x)$, loosely speaking. This issue is related to the variant of optimal transport problem that was the predecessor of Monge-Kantorovich problem – for pair of source and target measures we seek a *transport mapping* T being a Borel measurable function:

$$\inf_T \left\{ \int_{\overline{\Omega}} |T(x) - x|\,\nu_1(dx) \,:\, T\#\nu_1 = \nu_2 \right\}.$$

The problem above is known as the *Monge problem*. The infimum is not smaller than $W(\nu_1, \nu_2)$ and in general the solution does not exist. There are in fact simple examples of ν_1, ν_2 where there is no function T satisfying the constraint $T\#\nu_1 = \nu_2$; for more details the reader is referred to [166], [152]. By exploiting relation (4.75) we arrive at a simple consequence of Proposition 4.68:

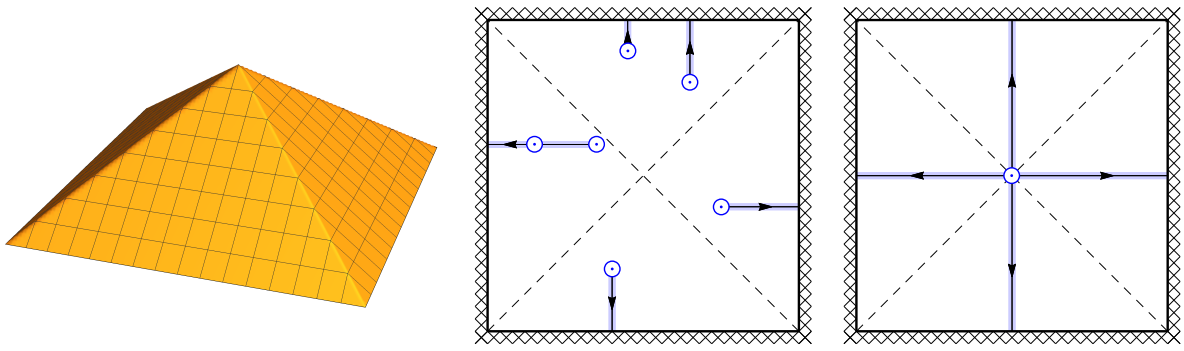
Corollary 4.70. *Let $\pi_{\partial\Omega} : \Omega \to \partial\Omega$ be any Borel measurable selection of $\Pi^{\partial\Omega}$. Then for any $f \in \mathrm{P}(\Omega)$ the triple*

$$\gamma = (\mathrm{id} \times \pi_{\partial\Omega})\#f, \qquad \nu_1^0 = 0, \qquad \nu_2^0 = \pi_{\partial\Omega}\#f$$

solves $(\mathrm{OTP}^{\partial\Omega})$ *for $\nu_1 = f$, $\nu_2 = 0$.*

Moreover, the mapping $T = \pi_{\partial\Omega}$ solves the Monge problem for $\nu_1 = f$ and $\nu_2 = \pi_{\partial\Omega}\#f$.

Example 4.71 (Transporting discrete mass to the boundary of a square). We consider a simple $(\mathrm{OTP}^{\partial\Omega})$ problem for a square domain $\Omega = (-a, a) \times (-a, a)$ and a discrete mass $f = \sum_{i=1}^n \alpha_i \, \delta_{x_i}$ where $x_i \in \Omega$, $\alpha_i \geq 0$ and $\sum_{i=1}^n \alpha_i = 1$ so that $f \in \mathrm{P}(\Omega)$. By virtue of Proposition 4.68 the sandpile-like function $u = d^{\partial\Omega}$ solves problem $(\overline{\mathcal{P}})$ for such class of f, regardless of the particular mass distribution; the diagram of u is showed in Fig. 4.11(a).



(a) (b) (c)

Figure 4.11: Optimal transport of discrete masses to the boundary of a square domain: (a) the universal "sandpile solution" $u = d^{\partial\Omega}$ to problem $(\overline{\mathcal{P}})$ for any $f \in \mathrm{P}(\Omega)$; (b) the unique optimal transport plan γ for six point masses; (c) the symmetric optimal plan γ for a single mass at the centre.

Fig. 4.11(b) presents a solution γ for six equal discrete masses, i.e. $\sum_{i=1}^6 \frac{1}{6} \, \delta_{x_i}$. Since each of the masses lies within interior of one of the four triangles separated by dashed lines, for each x_i the set $\Pi^{\partial\Omega}(x_i) = \{\pi_{\partial\Omega}(x)\}$ is a singleton. As a result there exists only one γ satisfying the prerequisite of Proposition 4.68, namely

$$\gamma = (\mathrm{id} \times \pi_{\partial\Omega})\#f = \sum_{i=1}^{6} \frac{1}{6} \, \delta_{(x_i, \pi_{\partial\Omega}(x_i))}$$

is the unique solution of $(\mathrm{OTP}^{\partial\Omega})$, cf. Remark 4.69.

For contrast, we consider a single mass located at the centre of the square: $f = \delta_{(0,0)}$. Clearly $\Pi^{\partial\Omega}\big((0,0)\big) = \big\{y_j\big\}_{j=1}^4$ where y_j are the four centres of square's edges. As a result, any measure $\gamma = \sum_{j=1} \gamma_j \, \delta_{((0,0),y_j)}$ with $\gamma_j \geq 0$, $\sum_{j=1}^4 \gamma_j = 1$ is a solution of ($\mathrm{OTP}^{\partial\Omega}$). Fig. 4.11(c) shows the symmetric solution obtained for $\gamma_j = \frac{1}{4}$; it will turn out to be of special significance for the optimal membrane problem, cf. Section 5.6. $\triangle$

4.9. Isotropic Masonry Material Design – discussion

Out of the anisotropy settings that had been put forth in Chapter 3 two of them were not employed in the examples: analytical in Section 4.5.1 or numerical in Section 4.6.2. These are: the Canvas-like Material Design (CanMD) and the isotropic Masonry Material Design (MMD) which were constituted by the energy functions j_+ and j_-, respectively. In the analytical case the reason for this omission is as follows: in the summary (4.19) that defines the elasticity Framework #1 of the (FMD) problem the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$) is implied and functions j_+, j_- do not satisfy it, see Proposition 3.23 and the comment above it.

An anisotropy setting that does not meet the requirement ($\mathrm{A}_{\mathrm{ellip}}$) is not automatically disqualified since the (FMD) problem was posed for any energy function satisfying (A1)-(A4). Without ($\mathrm{A}_{\mathrm{ellip}}$), however, the use of the results hitherto obtained is limited. The construction of optimal λ from solution of ($\mathcal{P}^*$) developed in Section 4.2 remains true, although there is no guarantee that $Z < \infty$ and the opposite case makes the construction useless. On the other hand Section 4.4 on the optimality conditions is completely out of bounds as the assumption (B5) explicitly requires that ($\mathrm{A}_{\mathrm{ellip}}$) is satisfied. The matter of exploiting Section 4.3 – which furnishes a relaxation of problem ($\mathcal{P}$) and guarantees that $Z < \infty$ – is more delicate. The assumptions (B1)-(B3) therein do not necessitate ($\mathrm{A}_{\mathrm{ellip}}$) but the compactness condition (B3) is difficult to meet without it: in particular the proof of Proposition 4.29 strongly relies on ($\mathrm{A}_{\mathrm{ellip}}$).

We shall now quickly show how conditions (B1)-(B3) can be revived for the two overlooked anisotropy settings that enforce the sign of the stress. Let us concentrate on the isotropic Masonry Material Design setting since the other, Canvas-like Material Design setting will be heavily employed in Chapters 5 and 6. We thus put forward a modification of the elasticity Framework #1 in (4.19):

$$\begin{cases} U = \mathbb{R}^d, \quad \mathcal{U} = \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d), \\ |||\cdot||| = \|\cdot\|_{L^1(\Omega)}, \\ \mathscr{F}(u) = \int_\Omega \langle u, F\rangle \, d\mathcal{L}^d \quad \text{where } F \in L^\infty(\Omega; \mathbb{R}^d), \\ V = \mathcal{S}^{d\times d}, \quad Au = e(u) := \frac{1}{2}\Big(\nabla u + (\nabla u)^\top\Big), \\ j_- : \mathcal{K} \times \mathcal{S}^{d\times d} \to \mathbb{R}_+ \quad \text{as in Example 3.27.} \end{cases}$$

The choice of a weaker norm $\|\cdot\|_{L^1(\Omega)}$ (than $\|\cdot\|_{\infty,\Omega}$ as was chosen in (4.19)) increases the chance for compactness (B3). However, it brutally limits the class of loads $\mathscr{F}$ that are continuous as required in (B2): it enforces that the load is represented by an L^∞ function

F, which in terms of mechanics allows *body forces* only. The severe restrictions do not end here. We will enforce that the region where the structure may rest surrounds the design domain Ω, i.e. we set the Dirichlet region $\Upsilon_0 = \partial\Omega$. Even then proving compactness is not easy, in fact the result below is a corollary from Theorem 5.25 that will be given in Section 5.3:

Proposition 4.72. *Let Ω be a bounded domain of Lipschitz boundary. Then, for ρ_- as in Example 3.27 and for $\Upsilon_0 = \partial\Omega$ the set*

$$\mathcal{U}_1 = \Big\{ u \in \mathcal{D}(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d) \,:\, \|\rho_-\big(e(u)\big)\|_{\infty,\Omega} \le 1 \Big\} \tag{4.76}$$

is pre-compact in $L^1(\Omega;\mathbb{R}^d)$. Its closure is a subset of $BV(\Omega;\mathbb{R}^d)$.

Proof. The set $-\mathcal{U}_1$ is exactly the projection onto the second coordinate of the set $\mathcal{X}_{1,\rho_+}$ defined in Section 5.3 for ρ_+ induced by the set $\mathcal{K} = \mathcal{K}_{iso}$. The assertion then follows by Theorem 5.25. □

The assumptions on the load F and the supporting region Υ_0 are by no means satisfactory in terms of practical design – especially $\Upsilon_0 = \partial\Omega$ almost never occurs in practice. After all it does not hurt to run the finite element simulation for a more realistic setting. First, in the spirit of Section 4.6.2, we must derive the conic quadratic formulation for the MMD setting. This can be done by a simple modification of the formulation proposed for the IMD setting, cf. Section 4.6.2.2. We make use of the equivalence that follows from Proposition 3.26: $\rho_-(\xi) \le 1$ if and only if $\rho(\xi+\zeta) \le 1$ for some $\zeta \in \mathcal{S}^{2\times 2}_-$. Therefore $\zeta \in \mathcal{S}^{2\times 2}_-$ may be added as the extra variable in the primal maximization problem. Without further explanation let us put forward the *conic quadratic finite element formulation for the isotropic Masonry Material Design*: the primal problem

$$\sup_{\substack{\mathbf{u}_1,\,\mathbf{u}_2\in\mathbb{R}^m\\ \mathbf{t},\mathbf{d},\boldsymbol{\epsilon}^{12}\in\mathbb{R}^n,\\ \mathbf{r}\in\mathbb{R}^n\\ \mathbf{t}_-,\mathbf{d}_-,\boldsymbol{\epsilon}^{12}_-\in\mathbb{R}^n}} \left\{ \langle \mathbf{u}_1, \mathbf{F}_1\rangle + \langle \mathbf{u}_2, \mathbf{F}_2\rangle \;:\; \begin{array}{cc} \mathbf{T}_1\mathbf{u}_1 + \mathbf{T}_2\mathbf{u}_2 + \mathbf{t}_- - \mathbf{t} = \mathbf{0} & \\ \mathbf{D}_1\mathbf{u}_1 + \mathbf{D}_2\mathbf{u}_2 + \mathbf{d}_- - \mathbf{d} = \mathbf{0} & \\ \mathbf{B}^{12}_1\mathbf{u}_1 + \mathbf{B}^{12}_2\mathbf{u}_2 + \boldsymbol{\epsilon}^{12}_- - \boldsymbol{\epsilon}^{12} = \mathbf{0} & \\ \mathbf{r} = \sqrt{2}\,\mathbf{1} & \\ -\mathbf{1} \le \mathbf{t} \le \mathbf{1} & \\ \big(\mathbf{r}(i), \mathbf{d}(i), \boldsymbol{\epsilon}^{12}(i)\big) \in \mathcal{K}^3_{\text{quad}} & \forall\, i \\ \big(-\mathbf{t}_-, \mathbf{d}_-(i), \boldsymbol{\epsilon}^{12}_-(i)\big) \in \mathcal{K}^3_{\text{quad}} & \forall\, i \end{array} \right\}$$

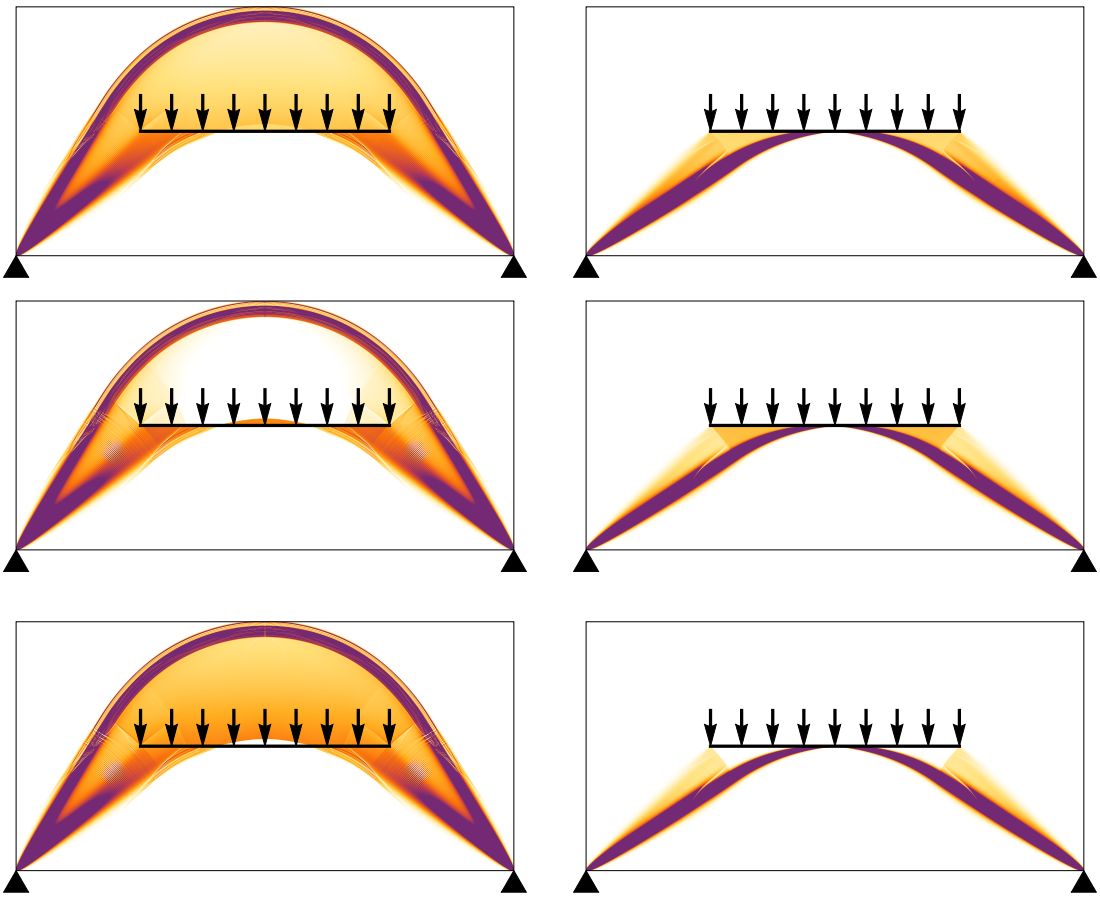
and its dual

$$\inf_{\substack{\mathbf{t}^*_-\in\mathbb{R}^n_+\\ \mathbf{d}^*,\boldsymbol{\sigma}^{12}\in\mathbb{R}^n,\\ \mathbf{r}^0\in\mathbb{R}^n}} \left\{ \big\langle \mathbf{1}, \mathbf{t}^*_-\big\rangle + \sqrt{2}\,\big\langle \mathbf{1}, \mathbf{r}^0\big\rangle \;:\; \begin{array}{cc} \mathbf{T}_1^\top(-\mathbf{t}^*_-) + \mathbf{D}_1^\top\mathbf{d}^* + (\mathbf{B}^{12}_1)^\top\boldsymbol{\sigma}^{12} = \mathbf{F}_1 & \\ \mathbf{T}_2^\top(-\mathbf{t}^*_-) + \mathbf{D}_2^\top\mathbf{d}^* + (\mathbf{B}^{12}_2)^\top\boldsymbol{\sigma}^{12} = \mathbf{F}_2 & \\ \big(\mathbf{r}^0(i), \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\big) \in \mathcal{K}^3_{\text{quad}} & \forall\, i \\ \big(\mathbf{t}^*_-, \mathbf{d}^*(i), \boldsymbol{\sigma}^{12}(i)\big) \in \mathcal{K}^3_{\text{quad}} & \forall\, i \end{array} \right\}$$

We shall now put the FE formulation above to use on an example Ω, Υ_0, F. Although we were dissatisfied with the limitation on Υ_0, F above we cannot choose them completely freely. Accordingly, for both the cantilever and the bridge problems previously investigated one

can prove that $Z = \infty$ for the MMD setting: in both cases there exists no pure-compression stress $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{2\times 2}_{-})$ such that $-\mathrm{Div}\,\varsigma = F$ in $\mathbb{R}^d \setminus \Upsilon_0$. In the case of the bridge problem this is easily remedied: it is enough to slightly (by any $\varepsilon > 0$) elevate the load F above the level of the supports Υ_0. To make the results more interesting we will additionally distribute the load along a horizontal segment. The simulation for MMD will be compared to the corresponding one for the Isotropic Material Design (IMD):

Example 4.73 (IMD vs the isotropic Masonry Material Design (MMD)). Let us take a rectangular domain $\Omega = (-1,1) \times (0,1)$, a two-point Dirichlet region $\Upsilon_0 = \big\{(-1,0),(1,0)\big\}$ and a vertical line load $F = -\mathrm{e}_2\,\mathcal{H}^1 \llcorner [a,b]$ where $a = (-\frac{1}{2},\frac{1}{2})$, $b = (\frac{1}{2},\frac{1}{2})$. The numerical simulations were carried out for the mesh counting 2 250 000 finite elements. Fig. 4.12 shows the results in the form of heat maps of $\mathrm{Tr}\mathscr{H}_h$, bulk modulus $\mathscr{K}_h$ and shear modulus $\mathscr{G}_h$ for two anisotropy settings of the problem (FMD_h): (a) the Isotropic Material Design setting (IMD) and (b) the isotropic Masonry Material Design setting (MMD). The respective values of Z_h are: $Z_h^{\mathrm{IMD}} = 2.8479$ and $Z_h^{\mathrm{MMD}} = 3.3287$.



(a) (b)

Figure 4.12: A modified bridge problem for an elevated line load – distribution of (from top to bottom) $\mathrm{Tr}\mathscr{H}_h$, bulk modulus $\mathscr{K}_h$, shear modulus $\mathscr{G}_h$ that are optimal in (FMD_h) for: (a) the Isotropic Material Design setting (IMD); (b) the isotropic Masonry Material Design setting (MMD).

The masonry bridge showed in Fig. 4.12(b) transfers the load to the supports by means of compression only. The isotropic structure in Fig. 4.12(a) additionally exploits tension: by means of it the load "hangs" on the arch in compression that constitutes the top edge of the structure. We can observe that in the case of the MMD setting the support of the function $\mathrm{Tr}\mathscr{H}_h$ – and automatically of the function σ_h – is contained in the convex hull of the set $\mathrm{spt}\, F \cup \Upsilon_0$. This adheres to the result established in the work [45], see Theorem 5.115 in Section 5.7.3. △

To sum up, if we wish to comprehend the theory of the Free Material Design in the case of the isotropic Masonry Material Design setting, a different strategy must be chosen that the one consisting in the relaxation procedure from Section 4.3. The work on masonry structures [90] or the one on cable webs in tension [45] may possibly serve as the starting point. This topic is scheduled for the future research of the present author.

Chapter 5

Optimal design of plane membranes

5.1. The plastic design of the pre-stress field in membranes (the linear model)

The simplest mathematical model of a plane membrane is similar to the one of a stationary heat conduction problem. Consequently, to some extent we shall follow Section 4.5.3 on optimal design of conductors. In principal membranes are two dimensional bodies that lie within a domain Ω contained in a horizontal plane $\mathbb{R}^2$. Notwithstanding this, the mathematics put forth in this chapter will be valid for the ambient space of arbitrary dimension d therefore we will consider bounded domains $\Omega \subset \mathbb{R}^d$. Membranes are subject to an out-of-plane (vertical, transverse) load that may be identified with a scalar field $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$.

In contrast to other mechanical models discussed in this dissertation, the stiffness of the membrane against f does not come from its elastic material's properties, at least not directly. Instead it is induced by the in-plane tensile stress $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$. In the linear model of a membrane this stress is introduced prior to loading with f, typically by exerting an in-plane load on the boundary $\partial\Omega$. As a result ς deserves to be called the *pre-stress field*. The condition on positive semi-definiteness of ς will be discussed in Section 5.2.1. Then, compliance under the transverse load f of such a membrane is given by

$$\begin{aligned} \mathcal{C}(\varsigma) &= \sup\left\{ \int u\, df - \int j_2(\varsigma, \nabla u) \ : \ u \in \mathcal{D}(\Omega) \right\} \\ &= \sup\left\{ \int u\, df - \frac{1}{2}\int \langle \nabla u \otimes \nabla u, \varsigma \rangle \ : \ u \in \mathcal{D}(\Omega) \right\} \end{aligned} \tag{5.1}$$

where the energy integrand $j_2 : \mathcal{S}^{d\times d}_+ \times \mathbb{R}^d \to \mathbb{R}_+$ reads

$$j_2(\sigma, \theta) := \frac{1}{2}\,\langle \sigma\theta, \theta \rangle.$$

It falls into the class of integrands j considered in Chapter 3, although we reinterpret the parameters: $H \in \mathcal{K}$ is now $\sigma \in \mathcal{S}^{d\times d}_+$ and $\xi \in V$ is $\theta \in \mathbb{R}^d$.

Above $u : \overline{\Omega} \to \mathbb{R}$ plays the role of an out-of-plane displacement function or a *deflection* function; condition $\operatorname{spt} u \subset \Omega$ essentially enforces that the membrane is pinned on $\partial\Omega$ in the transverse direction. One can readily see that the underlying mathematical model is identical to the one of a stationary heat conduction problem for $\Upsilon_0 = \partial\Omega$ (cf. (4.37) with

$j = j_2$) yet the conductivity field $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ is now translated to the pre-stress ς. Classically the pre-stress in a membrane is assumed to be homogeneous and hydrostatic (i.e. isotropic), namely $\varsigma = s\,\mathrm{I}\,\mathcal{L}^d \llcorner \Omega$ for $s \in \mathbb{R}_+$. In this case it is well established that the Euler-Lagrange equation for (5.1) is the Poisson equation $-s\,\triangle u = f$, see e.g. [61, Section IV.10.3].

In the spirit of the Free Material Design problem studied in Chapter 4, our goal will be to find the pre-stress ς that minimizes compliance $\mathcal{C}(\varsigma)$. We will impose the upper bound on $\int \operatorname{Tr}\varsigma$, although the field ς does not describe the material distribution *a priori*, therefore such a constraint is less natural than $\int c(\lambda) \le C_0$ in case of (FMD). For justification we may take an approach of *plastic design* that was recalled in Section 4.7.2 for Michell problem: the structural volume is then expressed by $\int c_{\mathrm{M}}(\varsigma)$ where c_{M} was given in (4.62). For $\sigma \in \mathcal{S}^{d\times d}_+$ we simply have $c_{\mathrm{M}}(\sigma) = \frac{1}{\bar{s}_+} \operatorname{Tr}\sigma$ where $\bar{s}_+ =: \bar{s}$ is the tensile yield stress. In the end the constraint $\frac{1}{\bar{s}} \int \operatorname{Tr}\varsigma \le V_0$ is a mechanically justified condition on the volume of the membrane.

Finally we add the constraint that will truly distinguish designing a membrane from designing a heat conductor. Since ς is a stress it must satisfy the equilibrium equations. Apart from the boundary $\partial\Omega$, where an in-plane *pre-loading* F is applied (cf. Section 5.7.5), in the interior Ω the only non-zero component of the load is the one that is out of the plane, i.e. f. The in-plane equilibrium of the membrane thus implies the *zero divergence constraint*:

$$\operatorname{Div}\varsigma = 0 \qquad \text{in } \Omega.$$

where we recall that the equation is intended in the sense of distributions in $\mathcal{D}'(\Omega; \mathbb{R}^d)$. We can readily pose the problem of designing the *Optimal Plastic Membrane*:

$$\mathcal{C}_{\min,0} = \inf\left\{ \mathcal{C}(\varsigma) \;:\; \varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+),\ \int \operatorname{Tr}\varsigma \le \bar{s}\,V_0,\ \ \operatorname{Div}\varsigma = 0 \ \text{ in } \Omega \right\} \qquad \text{(OPM)}$$

where "0" in the symbol $\mathcal{C}_{\min,0}$ ought to reflect the zero divergence constraint. This constraint rules out many competitors ς, for instance:

- a hydrostatic tensor field of the form $\varsigma = s\,\mathrm{I}\,\mathcal{L}^2 \llcorner \Omega$ is not admissible for (OPM) unless s is a constant;
- if $\varsigma = s(x)\,\tau(x) \otimes \tau(x)\,\mathcal{H}^1 \llcorner \mathrm{C}$ with C being a simple curve, s a positive weight and $\tau(x)$ a unit vector vector, then ς is admissible for (OPM) if and only if C is a straight line connecting two points of $\partial\Omega$ while τ is a constant vector parallel to C and s is a constant.

As a result, if we assume that $C_0 = \bar{s}\,V_0$, then clearly

$$\mathcal{C}_{\min,0} \ge \mathcal{C}_{\min} \tag{5.2}$$

where $\mathcal{C}_{\min}$ is the value function in the respective (FMD) problem.

To get a better picture of the difference between problems (FMD) and (OPM) let us look into the case of a one point source/load $f = \delta_{x_0}$ in a square domain Ω. We assume that $\Upsilon_0 = \partial\Omega$ in the (FMD) problem, therefore the optimal design of a heat conductor reduces to transporting the mass δ_{x_0} to the boundary with a minimal cost, see Section



(a) (b)

Figure 5.1: A one-point source/load $f = \delta_{x_0}$ in a square domain Ω: (a) solution $\check{\lambda}$ to (FMD) for a heat conductor with $\Upsilon_0 = \partial\Omega$; (b) solution $\check{\varsigma}$ to the optimal membrane problem (OPM). The points a_i are centres of the square's edges.

4.8.2 and Fig. 4.11(b). Hence the optimal conductivity field $\check{\lambda} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ concentrates on a single segment connecting x_0 to its projection onto the nearest edge of the square, cf. Fig. 5.1(a). We easily check that $-\mathrm{Div}\,\check{\lambda} = v\,\delta_{x_0}$ (computed in $\mathcal{D}'(\Omega;\mathbb{R}^d)$ where $v \in \mathbb{R}^2$ is a vector pointing to the left in the Fig. 5.1(a). We infer that solution $\check{\varsigma}$ of (OPM) must be different than $\check{\lambda}$: as it will unfold, solution $\check{\varsigma}$ concentrates on four segments $[x_0, a_i]$ connecting the load application point to the centres of the square's edges, cf. Fig. 5.1(b). The optimal membrane thus degenerates to the system of four strings that are tied at x_0. Since the strings are pre-stressed and pinned on $\partial\Omega$ they resist the vertical force f, i.e. the deflection of the point x_0 is finite. A proof of optimality of this membrane can be found in Example 5.62 below.

The zero divergence constraint corrupts the passage to the pair of problems $(\mathcal{P})$, $(\mathcal{P}^*)$ exploited in Chapter 4 for the (FMD) problem. For instance Lemma 4.2, that linked the values $\mathcal{C}_{\min}$ and $\sup\mathcal{P}$, relied on Proposition 3.39 whose proof consisted in choosing admissible fields of the form $\bar{\lambda} = \bar{H}\,\delta_{\bar{x}}$ whose divergence is clearly non-zero. Nonetheless we will manage to derive a new pair of mutually dual convex problems $(\mathscr{P})$, $(\mathscr{P}^*)$ upon whose solutions we will be able to reconstruct the optimal membrane. In this subsection, however, we will merely outline the direction of the reasoning, while Section 5.2 shall include the specifics. We start with a slight alteration of the problem (OPM) in which we dispose of constants $\bar{s}, V_0$:

$$\widetilde{\mathcal{Z}} := \inf\left\{\mathcal{C}(\varsigma) + \int \mathrm{Tr}\,\varsigma \;:\; \varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+),\ \mathrm{Div}\,\varsigma = 0 \ \text{ in } \Omega\right\}. \tag{5.3}$$

The two problems are related as follows:

Proposition 5.1. *The problems* (OPM) *and* (5.3) *admit solutions. Moreover, the following statements hold true:*

(i) the functional $\varsigma \mapsto \frac{1}{\mathcal{C}(\varsigma)}$ *is positively one-homogeneous, more precisely*

$$\mathcal{C}(t\,\varsigma) = \frac{1}{t}\,\mathcal{C}(\varsigma) \qquad \forall\, t > 0;$$

(ii) every solution $\hat\varsigma$ to the problem (5.3) *satisfies the equi-repartition of energy rule:*

$$\mathcal{C}(\hat\varsigma) = \int \operatorname{Tr}\hat\varsigma = \frac{\widetilde{\mathcal{Z}}}{2};$$

(iii) measure $\hat\varsigma \in \mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)$ solves (5.3) *if and only if $\check\varsigma = \frac{2\bar s V_0}{\widetilde{\mathcal{Z}}}\,\hat\varsigma$ solves* (OPM) *while*

$$\mathcal{C}_{\min,0} = \frac{(\widetilde{\mathcal{Z}})^2}{4\,\bar s\,V_0}.$$

Proof. Assertion (i) follows by substituting $u = \frac{1}{t}\,\tilde u$ in the supremum defining $\mathcal{C}(t\,\varsigma)$. Claim (ii) is due to the fact that the function $t \mapsto \mathcal{C}(t\hat\varsigma) + t\int \operatorname{Tr}\hat\varsigma = \frac{1}{t}\,\mathcal{C}(\hat\varsigma) + t\int \operatorname{Tr}\hat\varsigma$ is minimal at $t = 1$ (note that the constraints in (5.3) are positively homogeneous).

Let $\hat\varsigma$ be a solution to (5.3). Due to assertions (i) and (ii) we have $\int \operatorname{Tr}\check\varsigma = \frac{2\bar s V_0}{\widetilde{\mathcal{Z}}}\int \operatorname{Tr}\hat\varsigma = \bar s\,V_0$ while $\mathcal{C}(\check\varsigma) = \frac{\widetilde{\mathcal{Z}}}{2\bar s V_0}\,\mathcal{C}(\hat\varsigma) = \frac{(\widetilde{\mathcal{Z}})^2}{4\bar s V_0}$ therefore we infer that $\check\varsigma$ is admissible for (OPM) and thus $\mathcal{C}_{\min,0} \le \frac{(\widetilde{\mathcal{Z}})^2}{4\bar s V_0}$. Meanwhile, for any admissible ς for (OPM) the measure $\tilde\varsigma = \frac{\widetilde{\mathcal{Z}}}{2\bar s V_0}\,\varsigma$ yields $\int \operatorname{Tr}\tilde\varsigma \le \frac{\widetilde{\mathcal{Z}}}{2}$ therefore $\mathcal{C}(\tilde\varsigma) \ge \frac{\widetilde{\mathcal{Z}}}{2}$ by assertion (ii); as a result $\mathcal{C}(\varsigma) = \frac{\widetilde{\mathcal{Z}}}{2\bar s V_0}\,\mathcal{C}(\tilde\varsigma) \ge \frac{(\widetilde{\mathcal{Z}})^2}{4\bar s V_0}$ and by taking the infimum we discover the converse inequality $\mathcal{C}_{\min,0} \ge \frac{(\widetilde{\mathcal{Z}})^2}{4\bar s V_0}$ which renders $\check\varsigma$ a solution to (OPM). Conversely, if $\varsigma = \check\varsigma$ solves (OPM) then $\mathcal{C}(\hat\varsigma) = \mathcal{C}(\tilde\varsigma) = \frac{\widetilde{\mathcal{Z}}}{2}$ while $\int \operatorname{Tr}\hat\varsigma = \int \operatorname{Tr}\tilde\varsigma \le \frac{\widetilde{\mathcal{Z}}}{2}$, which ends the proof of (iii).

The proof of existence of solution to (OPM) employs the same arguments as in the proof of Proposition 4.1 and the fact that the set $\big\{\varsigma \in \mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}) : \operatorname{Div}\varsigma = 0 \text{ in } \Omega\big\}$ is weakly-* closed. The existence result for (5.3) is now a consequence of the assertion (iii) (cf. Proposition 5.14 for an independent proof). □

In the next step we will dispose of the zero divergence constraint in (5.3) by introducing the corresponding Lagrange multiplier w: by definition the equation $\operatorname{Div}\varsigma = 0$ in $\mathcal{D}'(\Omega;\mathbb{R}^d)$ holds if and only if for any $w \in \mathcal{D}(\Omega;\mathbb{R}^d)$ we have $\int \langle e(w),\varsigma\rangle = 0$. In addition we plug the definition (5.1) of $\mathcal{C}(\varsigma)$ to find that (note that $\operatorname{Tr}\varsigma = \langle \mathrm{I},\varsigma\rangle$):

$$\begin{aligned}
\widetilde{\mathcal{Z}} &= \inf_{\substack{\varsigma\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)\\ \operatorname{Div}\varsigma=0 \text{ in } \Omega}} \sup_{u\in\mathcal{D}(\Omega)} \left\{\int u\,df - \frac{1}{2}\int \langle \nabla u\otimes\nabla u,\varsigma\rangle + \int \operatorname{Tr}\varsigma\right\}\\
&= \inf_{\varsigma\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)} \sup_{\substack{u\in\mathcal{D}(\Omega)\\ w\in\mathcal{D}(\Omega;\mathbb{R}^d)}} \left\{\int u\,df - \frac{1}{2}\int \langle \nabla u\otimes\nabla u,\varsigma\rangle + \int \langle \mathrm{I},\varsigma\rangle - \int \langle e(w),\varsigma\rangle\right\}\\
&\le \sup_{\substack{u\in\mathcal{D}(\Omega)\\ w\in\mathcal{D}(\Omega;\mathbb{R}^d)}} \inf_{\varsigma\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)} \left\{\int u\,df + \int \left\langle \mathrm{I} - \frac{1}{2}\,\nabla u\otimes\nabla u - e(w),\varsigma\right\rangle\right\}.
\end{aligned}$$

Due to the lack of compactness the Ky Fan's result (cf. Theorem 2.5) cannot be employed to switch inf and sup with an equality. In the last line above the stress field ς is free of the zero divergence constraint and we may again choose it in the form $\varsigma = \bar\sigma\,\delta_{\bar x}$ for $\bar\sigma \in \mathcal{S}^{d\times d}_+$, $\bar x \in \overline{\Omega}$. Now assume that for a fixed pair (u,w) there is a point $\bar x \in \overline{\Omega}$ where

the matrix $\bar{\xi} := \big(\mathrm{I} - \frac{1}{2}\nabla u \otimes \nabla u - e(w)\big)(\bar{x})$ is not an element of $\mathcal{S}^{d\times d}_+$. Then clearly $\inf_{\bar{\sigma}\in\mathcal{S}^{d\times d}_+}\langle\bar{\xi},\bar{\sigma}\rangle = -\infty$, therefore the last line above can be rewritten to arrive at:

$$\widetilde{\mathcal{Z}} \leq \sup_{\substack{u\in\mathcal{D}(\Omega)\\ w\in\mathcal{D}(\Omega;\mathbb{R}^d)}} \left\{ \int u\,df \ : \ \frac{1}{2}\nabla u\otimes\nabla u + e(w) \preceq \mathrm{I} \ \text{ in } \overline{\Omega} \right\} \tag{5.4}$$

where for two matrices $A, B \in \mathcal{S}^{d\times d}$ we write $A \preceq B$ if and only if $B - A \in \mathcal{S}^{d\times d}_+$, being equivalent to: $\langle A, \tau\otimes\tau\rangle \leq \langle B, \tau\otimes\tau\rangle$ for every $\tau \in S^{d-1}$. Consequently, the constraint in (5.4) may be rewritten as

$$\max_{i\in\{1,\ldots,d\}} \lambda_i\left(\frac{1}{2}\nabla u\otimes\nabla u + e(w)\right) \leq 1 \qquad \text{in } \ \overline{\Omega}. \tag{5.5}$$

The problem in (5.4) can be compared with the problem $(\mathcal{P})$ specified for the heat conductor design, we recall that $Z = \sup\mathcal{P} = \sup_{u\in\mathcal{D}(\Omega)}\big\{\int u\,df : |\nabla u| \leq 1 \ \text{ in } \overline{\Omega}\big\}$. For $w = 0$ the constraint (5.5) reduces to $|\nabla u| \leq \sqrt{2}$ thus we deduce that $\sup(5.4) \geq \sqrt{2}\,Z$.

In the forthcoming sections it will appear that for the optimal design of membranes the problem (5.4), hereinafter denoted by $(\mathscr{P})$, will play the role analogous to the one played by $(\mathcal{P})$ for the (FMD) problem. The problem (5.3) will be further modified towards a problem $(\mathscr{P}^*)$ dual to $(\mathscr{P})$, while equality in (5.4) will be proved to always hold true.

The results presented in this chapter are an effect of collaboration between the author of this thesis and Guy Bouchitté, see the preprint [29].

5.2. Optimal design of elastic membranes (the non-linear model)

It is fair to deem the proposal of problem (OPM) of designing an optimal plastic membrane rather heuristic. First of all, one may question whether the pre-stress field ς in whole $\overline{\Omega}$ is in full control of the designer. Indeed, in practice we can exert an in-plane pre-load F on the boundary $\partial\Omega$ only, while the pre-stress field in Ω ends up being an elastic response of the membrane. The plastic approach where the stress ς is being designed directly is known from Michell problem, yet problem (OPM) is different as it also involves the kinematics of the membrane through $\mathcal{C}(\varsigma)$. Moreover, for the Michell problem we have at our disposal equivalent elastic formulations, one of which was put forward in Section 4.7.2 herein.

This section starts with a proposal of an elastic membrane model. Its inspiration comes from the operator $(u, w) \mapsto \frac{1}{2}\nabla u\otimes\nabla u + e(w)$ spontaneously emerging in (5.4). With w being the vector function of in-plane displacement this operator is the core ingredient of the plate model of von Kármán, see e.g. [60] or [113, Section 4.1]. It is well established that, for a homogeneous isotropic 2D body, an asymptotic analysis of the latter model leads to recovery of the linear membrane model discussed in Section 5.1, see for instance [57], [85, Section 6]. With a non-linear membrane model that builds upon the von Kármán model we will be able to pose an optimal design formulation in the spirit of the (FMD) problem. Now, however, the framework will be similar to the one of 2D elasticity (cf. Section 4.5.1), i.e. we will be looking for an optimal distribution of 4th order Hooke tensor

$\lambda \in \mathcal{M}(\overline{\Omega}; \mathscr{L}_+(\mathcal{S}^{d\times d}))$. A particular case of this formulation will be proved to be equivalent to the problem of optimal plastic membrane in Section 5.1.

5.2.1. The von Kármán-type model of an elastic membrane

The elastic membrane will be occupying a plane domain $\Omega \subset \mathbb{R}^d$, namely the physical context requires that $d = 2$ (we shall keep the symbol d nevertheless). The membrane will be composed of non-homogeneous anisotropic elastic material and so we will employ the theory put forth in Chapter 3. This involves choosing a triple $\mathcal{K}, j, c$ tailored for the setting of d-dimensional elasticity ($W = \mathscr{L}_+(\mathcal{S}^{d\times d})$, $V = \mathcal{S}^{d\times d}$), namely we assume:

- a set of admissible "Hooke tensors" $\mathcal{K}$ being a closed convex cone in the space $\mathscr{L}_+(\mathcal{S}^{d\times d})$ (consisting of positive semi-definite symmetric linear operators from $\mathcal{S}^{d\times d}$ to $\mathcal{S}^{d\times d}$);
- an energy function $j : \mathcal{K} \times \mathcal{S}^{d\times d} \to \mathbb{R}_+$ satisfying (A1)-(A4) (cf. Section 3.1) and moreover the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$) (cf. Section 3.2);
- a cost function $c : \mathcal{K} \to \mathbb{R}_+$ being the restriction of a norm on $\mathscr{L}(\mathcal{S}^{d\times d})$.

Henceforward $u : \overline{\Omega} \to \mathbb{R}$ and $w : \overline{\Omega} \to \mathbb{R}^d = \mathbb{R}^2$ will stand for the functions describing the membrane's displacements: the out-of-plane displacement u (the deflection) and the in-plane displacement w. Let us define the non-linear operator that for a pair $(u, w) \in C^1(\overline{\Omega}) \times C^1(\overline{\Omega}; \mathbb{R}^d)$ returns a function in $C(\overline{\Omega}; \mathcal{S}^{d\times d})$:

$$\mathcal{A}(u, w) := \frac{1}{2} \nabla u \otimes \nabla u + e(w).$$

This operator will be referred to as the *von Kármán operator* (not to be confused with the *von Kármán bracket*, see [129]).

We shall now briefly recall the main elements of the von Kármán plate model, see [113, Section 4.1] for details. Two types of loads are assumed: the in-plane load $F : \Omega \to \mathbb{R}^2$ and the out-of-plane load $f : \Omega \to \mathbb{R}$, therefore the load potential reads $\int u\, f + \int \langle w, F \rangle$. With h standing for the plate's thickness, the elastic energy splits into two terms, i.e. $J_{\mathrm{vK}}(u, w) = J_{\mathrm{vK},1}(u, w) + J_{\mathrm{vK},2}(u)$ where $J_{\mathrm{vK},1}(u, w) = h \int j\big(\lambda, \mathcal{A}(u, w)\big)$ and $J_{\mathrm{vK},2}(u) = \frac{h^3}{12} \int j\big(\lambda, \nabla^2 u\big)$, where $j(H, \xi) = \frac{1}{2} \langle H\xi, \xi \rangle$. The component $J_{\mathrm{vK},1}(u, w)$ relies on the assumption that the strain of the plate's middle surface $\epsilon : \Omega \to \mathcal{S}^{2\times 2}$ is equal to $\mathcal{A}(u, w)$, in contrast to the classical plate theories where $\epsilon = e(w)$. The additional non-linear term $\frac{1}{2} \nabla u \otimes \nabla u$ accounts for large transverse deformation of the plate and in fact is the essence of the von Kármán model. The second ingredient $J_{\mathrm{vK},2}(u)$ is the energy of the Kirchhoff plate (cf. Section 4.5.2) that incorporates the bending effects.

The membrane model that the we propose comes from the following modifications of the von Kármán plate model:

- the membrane is subject to the out-of-plane load $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ only, more precisely the in-plane loading is absent in Ω:

$$F = 0 \quad \text{in} \quad \Omega; \tag{5.6}$$

- the membrane has a small thickness h, which leads to assumptions:

- the elastic energy due to bending is negligible, i.e.

$$J(u,w) = J_{\mathrm{vK},1}(u,w), \tag{5.7}$$

- the membrane is perfectly immune to local buckling, i.e. compression is excluded, which by Section 3.3.2 amounts to replacing the energy potential j by j_+, where

$$j_+(H,\,\cdot\,) = \big(j^*(H,\,\cdot\,) + \chi_{\mathcal{S}^{d\times d}_+}(\,\cdot\,)\big)^*. \tag{5.8}$$

Membranes are typically made of either very thin steel sheet or canvas, therefore assuming small thickness h is natural when comparing to concrete slabs, for instance. The fastest explanation behind the resulting assumption (5.7) is the following: we have $J_{\mathrm{vK},1}(u,w) \sim h$ while $J_{\mathrm{vK},2}(u) \sim h^3$ hence the bending energy tends to zero faster when $h \to 0$. This simple reasoning is often used to convert general theories of shells to membrane shell theories. The more sophisticated justification of the elastic energy depending on the strain $\mathcal{A}(u,w)$ only could possibly lead through an asymptotic analysis of a 3D body $\Omega \times (-h/2, h/2)$ with a suitably adjusted scaling (see e.g. [58] or [43]). This matter, however, lies beyond the scope of this thesis. Exclusion of compression (5.8) is particularly easy to defend if we assume that the membrane is made of canvas – then even small compressive in-plane stress leads to wrinkling of the canvas, which effectively reduces the membrane's resistance against compression to zero. In case of a membrane being a metal sheet the stiffness under compression is non-zero, however it is very small if the sheet is very thin due to the immunity to local buckling.

The elasticity problem of a membrane, whose assumptions were outlined and discussed above, may be summarized by defining its compliance. For an elastic material distribution given by a constitutive field $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ the compliance reads:

$$\mathscr{C}(\lambda) := \sup\left\{ \int u\,df - \int j_+\big(\lambda, \mathcal{A}(u,w)\big) \;:\; (u,w) \in \mathcal{D}\big(\Omega; \mathbb{R}^{d+1}\big) \right\}. \tag{5.9}$$

This definition entails the boundary conditions: at this point we assume that both the out-of-plane and the in-plane displacements u and w are zero on $\partial\Omega$. This framework rules out pre-stressing the membrane, e.g. by applying an in-plane boundary pre-loading – this possibility shall be discussed in Section 5.7.5. Nonetheless the in-plane stress $\sigma : \overline{\Omega} \to \mathcal{S}^{d\times d}$ will emerge as an elastic reaction to the out-of-plane load f. This phenomenon is precisely due to the non-linear term $\frac{1}{2}\,\nabla u \otimes \nabla u$ that enters the energy integrand j.

The mapping $(u,w) \mapsto J_{\mathrm{vK},1}(u,w)$ in the original von Kármán model, i.e. for $J_{\mathrm{vK},1}(u,w) = \int j\big(\lambda, \mathcal{A}(u,w)\big)$, is well known to be non-convex, which in fact enables an analysis of stability of plates, see [113, Section 4.3]. We shall now show that the change from j to j_+ restores the convexity, which is crucial in the perspective of optimal design. First, let us recall that:

Remark 5.2. Owing to Proposition 3.20 the function j_+ satisfies assumptions (A1)-(A4) from Section 3.1, but does not satisfy the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$). As a consequence, all the results from Chapter 3 – that did not require ($\mathrm{A}_{\mathrm{ellip}}$) – hold for $j = j_+$ and for the corresponding $\rho = \rho_+$.

We introduce a function $g : \mathcal{K} \times (\mathbb{R}^d \times \mathcal{S}^{d\times d}) \to \mathbb{R}_+$ as follows:

$$g\big(H,(\theta,\xi)\big) = j_+\Big(H, \tfrac{1}{2}\,\theta\otimes\theta + \xi\Big).$$

We will show that $g(H,\,\cdot\,)$ is convex for any $H \in \mathcal{K}$ and we will compute its convex conjugate. Beforehand we need to calculate the convex conjugate of the quadratic energy function $\mathbb{R}^d \ni \theta \mapsto j_2(\sigma,\theta)$ for given $\sigma \in \mathcal{S}^{d\times d}_+$, which in accordance with Chapter 3 we denote by $j_2^*(\sigma,\,\cdot\,)$. For any $q \in \mathbb{R}^d$ we obtain

$$j_2^*(\sigma,q) = \sup_{\theta\in\mathbb{R}^d}\big\{\langle\theta,q\rangle - \tfrac{1}{2}\,\langle\sigma\theta,\theta\rangle\big\} = \begin{cases} \frac{1}{2}\langle\sigma\hat\theta,\hat\theta\rangle & \text{if } \ q=\sigma\hat\theta \ \text{ for } \ \hat\theta\in\mathbb{R}^d,\\ \infty & \text{if } \ q\notin \mathrm{Im}(\sigma). \end{cases} \tag{5.10}$$

We note that in the case when $q \in \mathrm{Im}(\sigma)$ the value of $\frac{1}{2}\langle\sigma\hat\theta,\hat\theta\rangle$ is invariant of the choice of $\hat\theta$ since $q \perp \mathrm{Ker}(\sigma)$ due to symmetry of σ. If σ is strictly positive definite then we simply have $j_2^*(\sigma,q) = \frac{1}{2}\langle\sigma^{-1}q,q\rangle$. The latter formula is concise and informative for the reader, therefore we agree to abuse the notation and write

$$\tfrac{1}{2}\langle\sigma^{-1}q,q\rangle := j_2^*(\sigma,q)$$

while keeping in mind that $\frac{1}{2}\langle\sigma^{-1}q,q\rangle = \infty$ once $q \notin \mathrm{Im}(\sigma)$.

Lemma 5.3. *For any $H \in \mathcal{K}$ the function $\mathbb{R}^d \times \mathcal{S}^{d\times d} \ni (\theta,\xi) \mapsto g\big(H,(\theta,\xi)\big) \in \mathbb{R}_+$ is convex continuous. Its convex conjugate reads:*

$$\begin{aligned} g^*\big(H,(q,\sigma)\big) &= j_+^*(H,\sigma) + \tfrac{1}{2}\langle\sigma^{-1}q,q\rangle \\ &= \begin{cases} j^*(H,\sigma) + \frac{1}{2}\langle\sigma\hat\theta,\hat\theta\rangle & \text{if } \sigma\in\mathcal{S}^{d\times d}_+ \ \text{ and } \ q=\sigma\hat\theta \ \text{ for } \ \hat\theta\in\mathbb{R}^d,\\ \infty & \text{if } \sigma\notin\mathcal{S}^{d\times d}_+ \ \text{ or if } \ q\notin\mathrm{Im}(\sigma) \end{cases} \end{aligned} \tag{5.11}$$

while subdifferential of g with respect to the pair (θ,ξ) enjoys the characterization:

$$(q,\sigma) \in \partial g\big(H,(\theta,\xi)\big) \qquad \Leftrightarrow \qquad \begin{cases} \sigma \in \partial j_+\big(H, \frac{1}{2}\,\theta\otimes\theta+\xi\big),\\ q = \sigma\theta. \end{cases} \tag{5.12}$$

Proof. By definition of j_+ we have

$$\begin{aligned} g\big(H,(\theta,\xi)\big) = j_+\big(H, \tfrac{1}{2}\,\theta\otimes\theta+\xi\big) &= \sup_{\sigma\in\mathcal{S}^{d\times d}}\Big\{\big\langle\tfrac{1}{2}\,\theta\otimes\theta+\xi,\sigma\big\rangle - j^*(H,\sigma) - \chi_{\mathcal{S}^{d\times d}_+}(\sigma)\Big\}\\ &= \sup_{\sigma\in\mathcal{S}^{d\times d}_+}\big\{\big\langle\tfrac{1}{2}\,\theta\otimes\theta+\xi,\sigma\big\rangle - j^*(H,\sigma)\big\}\,. \end{aligned}$$

For any positive semi-definite matrix σ the function $(\theta,\xi) \mapsto \big\langle\frac{1}{2}\,\theta\otimes\theta+\xi,\sigma\big\rangle$ is convex continuous. Since the function $g(H,\,\cdot\,)$ is a point-wise supremum of a family of such functions it is convex and lower semi-continuous, cf. Proposition 2.8. In addition it is finite

(cf. assumption (A1)) on the whole space $\mathbb{R}^d \times \mathcal{S}^{d\times d}$ and therefore it is continuous as well. We compute its convex conjugate:

$$\begin{aligned} g^*\big(H,(q,\sigma)\big) &= \sup_{\theta\in\mathbb{R}^d}\sup_{\xi\in\mathcal{S}^{d\times d}}\big\{\langle\theta,q\rangle+\langle\xi,\sigma\rangle - j_+\big(H,\tfrac{1}{2}\,\theta\otimes\theta+\xi\big)\big\}\\ &= \sup_{\theta\in\mathbb{R}^d}\sup_{\tilde\xi\in\mathcal{S}^{d\times d}}\Big\{\langle\theta,q\rangle+\Big\langle\tilde\xi-\tfrac{1}{2}\,\theta\otimes\theta,\sigma\Big\rangle - j_+\big(H,\tilde\xi\big)\Big\}\\ &= \sup_{\theta\in\mathbb{R}^d}\Big\{\langle\theta,q\rangle - j_2(\sigma,\theta)\Big\} + \sup_{\tilde\xi\in\mathcal{S}^{d\times d}}\Big\{\langle\tilde\xi,\sigma\rangle - j_+\big(H,\tilde\xi\big)\Big\} = j_2^*(\sigma,q)+j_+^*(H,\sigma)\end{aligned}$$

where to pass to the second line we made a change of variables: $\tilde\xi=\frac{1}{2}\,\theta\otimes\theta+\xi$. The final formula for $g^*\big(H,(q,\sigma)\big)$ in (5.11) follows once we recall that $j_+^*(H,\sigma)=j(H,\sigma)$ when $\sigma\in\mathcal{S}^{d\times d}_+$ and $j_+^*(H,\sigma)=\infty$ otherwise. From the above we infer that the extremality equality $\langle\theta,q\rangle+\langle\xi,\sigma\rangle = g\big(H,(\theta,\xi)\big)+g^*\big(H,(q,\sigma)\big)$ holds true if and only if $\langle\theta,q\rangle = j_2(\sigma,\theta)+j_2^*(\sigma,q)$ and $\langle\tilde\xi,\sigma\rangle = j_+\big(H,\tilde\xi\big)+j_+^*\big(H,\tilde\sigma\big)$ (note that all the conjugates are computed with respect to the second argument). Characterization (5.12) follows. □

Since the operators ∇, e are linear, from the result above we immediately obtain convexity of the elastic energy integral, which distinguishes the proposed membrane model from the von Kármán plate:

Proposition 5.4. *For any $\lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K})$ the functional*

$$\mathcal{D}(\Omega;\mathbb{R}^{d+1})\ni(u,w)\mapsto\int j_+\big(\lambda,\mathcal{A}(u,w)\big)$$

is convex.

In order to derive the dual formula for the compliance $\mathscr{C}(\lambda)$ we require an additional technical result being a variation of Proposition 3.36 and Corollary 3.37:

Lemma 5.5. *For any measure $\mu\in\mathcal{M}_+(\overline{\Omega})$ and any measurable function $\mathscr{H}:\overline{\Omega}\to\mathcal{K}$ the map $g\big(\mathscr{H}(\,\cdot\,),\,\cdot\,\big):\overline{\Omega}\times(\mathbb{R}^d\times\mathcal{S}^{d\times d})\to\mathbb{R}_+$ is a Carathéodory function. If in addition $\mathscr{H}\in L^\infty_\mu(\overline{\Omega};\mathcal{K})$ then the convex functional*

$$L^{2p}_\mu(\overline{\Omega};\mathbb{R}^d)\times L^p_\mu(\overline{\Omega};\mathcal{S}^{d\times d})\ni(\phi,\epsilon)\mapsto\int g\big(\mathscr{H},(\phi,\epsilon)\big)\,d\mu$$

is continuous while its convex conjugate reads $L^{(2p)'}_\mu(\overline{\Omega};\mathbb{R}^d)\times L^{p'}_\mu(\overline{\Omega};\mathcal{S}^{d\times d})\ni(q,\sigma)\mapsto\int g^*\big(\mathscr{H},(q,\sigma)\big)\,d\mu$.

Proof. Upon acknowledging Lemma 5.3 the proof of the Carathéodory property is in full analogy to the proof of Proposition 3.36. The last statement relies on similar arguments to those behind Corollary 3.37 although the estimate used therein requires an additional step:

$$\begin{aligned}\int g\big(\mathscr{H},(\phi,\epsilon)\big)d\mu=\int j_+\big(\mathscr{H},\tfrac{1}{2}\,\phi\otimes\phi+\epsilon\big)d\mu &\le \|c(\mathscr{H})\|_{L^\infty_\mu}\frac{(C_1)^p}{p}\left(\big\|\tfrac{1}{2}\,\phi\otimes\phi+\epsilon\big\|_{L^p_\mu}\right)^p\\ &\le \|c(\mathscr{H})\|_{L^\infty_\mu}\frac{(\widetilde{C}_1)^p}{p}\left(\|\phi\|^{2p}_{L^{2p}_\mu}+\|\epsilon\|^p_{L^p_\mu}\right).\end{aligned}$$

□

Proposition 5.6. *For any representation $\lambda = \mathscr{H}\mu$ of $\lambda \in \mathcal{M}(\overline{\Omega};\mathcal{K})$ such that $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\mathscr{H} \in L^\infty_\mu(\overline{\Omega};\mathcal{K})$ there holds*

$$\mathscr{C}(\lambda) = \inf\Big\{ \int j^*(\mathscr{H},\sigma)\,d\mu + \int \tfrac{1}{2}\langle \sigma^{-1}q, q\rangle\,d\mu : \begin{array}{ll} q \in L^{(2p)'}_\mu(\overline{\Omega};\mathbb{R}^d), & -\mathrm{div}(q\mu) = f \\ \sigma \in L^{p'}_\mu(\overline{\Omega};\mathcal{S}^{d\times d}_+), & -\mathrm{Div}(\sigma\mu) = 0 \end{array} \text{ in } \Omega \Big\} \tag{5.13}$$

whereas the minimum is achieved whenever $\mathscr{C}(\lambda) < \infty$.

Proof. The assertion can be proved almost directly by employing Theorem 2.12. We define:

$$X = \Big\{(u,w) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1})\Big\}, \quad Y = L^{2p}_\mu(\overline{\Omega};\mathbb{R}^d) \times L^p_\mu(\overline{\Omega};\mathcal{S}^{d\times d}), \quad \Lambda(u,w) = \big(\nabla u, e(w)\big)$$

and $\Phi(u,w) = -\int u\,df$, $\Psi(\phi,\epsilon) = \int g\big(\mathscr{H},(\phi,\epsilon)\big)\,d\mu$. Then, for $(q,\sigma) \in Y^* = L^{(2p)'}_\mu(\overline{\Omega};\mathbb{R}^d) \times L^{p'}_\mu(\overline{\Omega};\mathcal{S}^{d\times d})$ we can check that $\Lambda^*(q,\sigma) = \big(-\mathrm{div}(q\mu), -\mathrm{Div}(\sigma\mu)\big)$ and that $\Phi^*\big(-\Lambda^*(q,\sigma)\big)$ is finite if and only if $-\mathrm{div}(q\mu) = f$ and $-\mathrm{Div}(\sigma\mu) = 0$. Directly by Lemma 5.5 we find that $\Psi^*(q,\sigma) = \int g^*\big(\mathscr{H},(q,\sigma)\big)\,d\mu$ and the claim follows by Theorem 2.12 and formula (5.11). Indeed: instead of writing the $\int j^*_+(\mathscr{H},\sigma)\,d\mu$ in the minimized functional while $\sigma \in L^{p'}_\mu(\overline{\Omega};\mathcal{S}^{d\times d})$, we may write $\int j^*(\mathscr{H},\sigma)\,d\mu$ if we require that $\sigma \in L^{p'}_\mu(\overline{\Omega};\mathcal{S}^{d\times d}_+)$. □

At this point our model of an elastic membrane is encompassed by two mutually dual variational problems (5.9) and (5.13) where (u,w) and, respectively, (q,σ) are the unknowns. We will now give the system of partial differential equations for the pairs (u,w), (q,σ) that are the criteria of optimality for the respective problems. Since $\sup(5.9) = \inf(5.13)$ then, apart from the admissibility conditions, we require that $\langle \nabla u, q\rangle + \langle e(w),\sigma\rangle = g\big(\mathscr{H},(\nabla u, e(w))\big) + g^*\big(\mathscr{H},(q,\sigma)\big)$ μ-a.e., cf. the formulas for g, g^* above and the arguments used in proof of Proposition 4.17. By exploiting the characterization (5.12) of the subdifferential of $g(H,\cdot)$ we infer that the pairs (u,w) and (q,σ) solve the problems (5.9) and (5.13) if and only if the following conditions are fulfilled:

$$\begin{cases} (i) & -\mathrm{div}(q\mu) = f, \quad \mathrm{Div}(\sigma\mu) = 0 \quad \text{in } \Omega, \\ (ii) & \sigma \in \partial j_+\big(\mathscr{H}, \tfrac{1}{2}\nabla u \otimes \nabla u + e(w)\big) \quad \mu\text{-a.e.}, \\ (iii) & q = \sigma\,\nabla u \quad \mu\text{-a.e.} \end{cases} \tag{5.14}$$

The derived system of PDE makes it possible to understand better the mechanical role of functions $q \in L^{(2p)'}_\mu(\overline{\Omega};\mathbb{R}^d)$ and $\sigma \in L^{p'}_\mu(\overline{\Omega};\mathcal{S}^{d\times d}_+)$. The function σ is the density of the absolute in-plane stress $\varsigma = \sigma\mu$ in the membrane. It is associated to the in-plane strain $\epsilon = \mathcal{A}(u,w) = \frac{1}{2}\nabla u \otimes \nabla u + e(w)$ through a non-linear constitutive Hooke's law point-wise induced by the potential $j_+\big(\mathscr{H}(x),\cdot\big)$ (it is non-linear even in the case when $j(H,\xi) = \frac{1}{2}\langle H\xi,\xi\rangle$ due to the "+" operation, cf. Section 3.3.2 for more details). The vector function q may be identified with density of the *transverse force field* which "carries" the vertical load f. The field q is more difficult to describe since it is not an independent stress as it is the case for some classical plate theories. Loosely speaking, $q(x)$ is the vertical component of the suitable stress resultants of $\sigma(x)$ that are rotated with respect to the base plane due to the slope $\nabla u(x)$.

The equations in (5.14) clearly require smoothness of the pair (u, w). At this moment relaxations of the maximization problem (5.9) is unclear. A partial success could be achieved if we allowed that $(u, w) \in \mathrm{dom}(\nabla_\mu) \times \mathrm{dom}(e_\mu)$, see a similar result in Proposition 4.17. However, in contrast to the Free Material Design problems studied in Chapter 4, in general solutions (u, w) must be sought outside those spaces, see Section 5.4 for more details.

5.2.2. Formulation of the optimal elastic membrane problem

In this short subsection we shall formulate the problem of designing an elastic membrane. The membrane model put forth above was tailored to keep the spirit of the Free Material Design covered in Chapter 4. The problem of *Optimal Elastic Membrane* may be readily proposed:

$$\mathscr{C}_{\min} = \min\left\{\mathscr{C}(\lambda) \ : \ \lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K}), \ \int c(\lambda) \leq C_0\right\} \tag{OEM}$$

It must be stressed that the (OEM) problem is not a special case of the (FMD) problem. In spite of plenty similarities the crucial difference is the following: the elastic energy functional $(u, w) \mapsto \int j_+\big(\lambda, \mathcal{A}(u, w)\big)$ is not positively p-homogeneous. More precisely it is $2p$-homogeneous and p-homogeneous in u and w respectively. Notwithstanding this discrepancy we will be able to develop similar methods to those employed in Chapter 4 and the fact that the in-plane load is zero (see (5.6)) will play a valid role.

At this stage already we may give the existence result for (OEM) and the argumentation will be identical as for (FMD):

Proposition 5.7. *There always exists a solution $\check{\lambda}$ of the* (OEM) *problem.*

Proof. For any $(u, w) \in \mathcal{D}(\Omega; \mathbb{R}^{d+1})$ the functional $\lambda \mapsto \int u \, df - \int j_+\big(\lambda, \mathcal{A}(u, w)\big)$ is convex and weakly-* lower semi-continuous on $\mathcal{M}(\overline{\Omega}; \mathscr{L}_+(\mathcal{S}^{d\times d}))$, see Theorem 3.35 and Remark 5.2. The rest of the proof is identical to the proof of Proposition 4.1. □

Remark 5.8. At this point the existence result should stipulate the case when $\mathscr{C}_{\min} = \infty$. However, in Section 5.3 we will show that $\mathscr{C}_{\min}$ is always finite thanks to assumption $(\mathrm{A}_{\mathrm{ellip}})$ on $\mathcal{K}, j$ (but not for $\mathcal{K}, j_+$).

5.2.3. The underlying pair of convex problems $(\mathscr{P})$ and $(\mathscr{P}^*)$

In Section 4.2 we have shown that the Free Material Design problem (FMD) could be reduced to a pair of convex problems $(\mathcal{P}), (\mathcal{P}^*)$. In a similar fashion here we will formulate a new pair of problems $(\mathscr{P}), (\mathscr{P}^*)$ that will play an analogous role for the Optimal Elastic Membrane problem (OEM). The route from (OEM) to $(\mathscr{P}), (\mathscr{P}^*)$ and back will not be identical, although it will share plenty of the same or similar steps. While making the proofs we will often confine ourselves to pointing out the differences with respect to the arguments from Chapter 4 to avoid repetitions.

We put forward the primal, displacement based problem:

$$\mathscr{Z} := \sup\left\{\int u \, df \ : \ (u, w) \in \mathcal{D}(\Omega; \mathbb{R}^{d+1}), \ \rho_+\left(\frac{1}{2}\nabla u \otimes \nabla u + e(w)\right) \leq 1 \ \text{ in } \overline{\Omega}\right\} \tag{$\mathscr{P}$}$$

We recall that $\rho_+ : \mathcal{S}^{d\times d} \to \mathbb{R}_+$ is the convex one-homogeneous lower semi-continuous function that, together with its polar ρ_+^0, is defined through the convex programs:

$$\frac{1}{p}\big(\rho_+(\xi)\big)^p = \max_{H\in\mathcal{K},\ c(H)\le 1} j_+(H,\xi), \qquad \frac{1}{p'}\big(\rho_+^0(\sigma)\big)^{p'} = \inf_{H\in\mathcal{K},\ c(H)\le 1} j_+^*(H,\sigma). \tag{5.15}$$

Based on Section 3.3.2 the following relations hold true:

$$\rho_+(\xi) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+} \rho(\xi+\zeta), \qquad \rho_+^0(\sigma) = \begin{cases} \rho^0(\sigma) & \text{if } \sigma\in\mathcal{S}^{d\times d}_+, \\ \infty & \text{if } \sigma\notin\mathcal{S}^{d\times d}_+ \end{cases} \tag{5.16}$$

where ρ, ρ^0 are computed for the initial energy function j, i.e.

$$\frac{1}{p}\big(\rho(\xi)\big)^p = \max_{H\in\mathcal{K},\ c(H)\le 1} j(H,\xi), \qquad \frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = \min_{H\in\mathcal{K},\ c(H)\le 1} j^*(H,\sigma). \tag{5.17}$$

The link between (OEM) and ($\mathscr{P}$) is as follows:

Lemma 5.9. *The minimum value of compliance in* (OEM) *equals*

$$\mathscr{C}_{\min} = \frac{2p-1}{2p}\left(\frac{\mathcal{Z}^{2p}}{2C_0}\right)^{\frac{1}{2p-1}}, \tag{5.18}$$

where $\mathcal{Z} = \sup\mathscr{P}$. *Moreover, with* $\check\lambda$ *being any solution of* (OEM) *the problem* (5.9) *for* $\lambda = \check\lambda$ *admits a maximizing sequence* $(u_n, w_n) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1})$ *that satisfies a uniform estimate:*

$$\|\rho_+\big(\mathcal{A}(u_n,w_n)\big)\|_{\infty,\Omega} \le \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{2}{2p-1}}. \tag{5.19}$$

Proof. By plugging definition (5.9) into (OEM) we arrive at a min-max problem. Thanks to Proposition 5.4 we have concavity of the mapping $(u,w)\mapsto \int u\,df - \int j_+\big(\lambda,\mathcal{A}(u,w)\big)$ and therefore, by repeating arguments from the proof of Lemma 4.2 we obtain:

$$\begin{aligned}\mathscr{C}_{\min} &= \inf_{\lambda\in\mathcal{M}(\overline\Omega;\mathcal{K}),\ \int c(\lambda)\le C_0}\ \sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{\int u\,df - \int j_+\big(\lambda,\mathcal{A}(u,w)\big)\right\} \\ &= \sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{\int u\,df - \frac{C_0}{p}\,\Big(\big\|\rho_+\big(\tfrac12\,\nabla u\otimes\nabla u + e(w)\big)\big\|_{\infty,\Omega}\Big)^p\right\}.\end{aligned}$$

The next step requires a variation of the technique used to prove Lemma 4.2: for $t\ge 0$ this time we substitute $u = t\,u_1$, $w = t^2 w_1$ in order to obtain

$$\begin{aligned}\mathscr{C}_{\min} &= \sup_{t\ge0}\ \sup_{(u_1,w_1)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{\left(\int u_1\,df\right)t - \frac{C_0}{p}\,t^{2p} :\ \big\|\rho_+\big(\tfrac12\,\nabla u_1\otimes\nabla u_1 + e(w_1)\big)\big\|_{\infty,\Omega}\le 1\right\} \\ &= \sup_{t\ge0}\left\{\mathcal{Z}\,t - \frac{C_0}{p}\,t^{2p}\right\} = \frac{2p-1}{2p}\left(\frac{\mathcal{Z}^{2p}}{2C_0}\right)^{\frac{1}{2p-1}}\end{aligned} \tag{5.20}$$

where the maximum with respect to t was found for $\bar t = \big(\frac{\mathcal{Z}}{2C_0}\big)^{\frac{1}{2p-1}}$. For $\check\lambda$ being a solution to (OEM) we repeat the final steps of the proof of Lemma 4.2 to find that

$$\mathscr{C}(\check\lambda) = \sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{\int u\,df - \int j_+\big(\lambda,\mathcal{A}(u,w)\big)\ :\ \|\rho_+\big(\mathcal{A}(u,w)\big)\|_{\infty,\Omega}\le \bar t^{\,2}\right\},$$

which furnishes the "moreover part" of the claim. □

Remark 5.10. The proof o Lemma 5.9 heavily relied on the assumption that the in-plane load F is zero, cf. (5.6). Otherwise the potential $\int u\,df$ would have always been accompanied by $\int \langle w, F\rangle$. As an effect the passage to (5.20) would have not been possible due to the difference in scaling: $u = t\,u_1$, $w = t^2 w_1$.

At the first glance the problem $(\mathscr{P})$ lies outside the class of problems encompassed by $(\mathcal{P})$: the operator $(u, w) \mapsto \mathcal{A}(u, w) = \frac{1}{2}\nabla u \otimes \nabla u + e(w)$ is non-linear. The idea now is to show that there is a convex positively one-homogeneous lower semi-continuous function $\varrho : \mathbb{R}^d \times \mathcal{S}^{d\times d} \to \mathbb{R}_+$ so that the condition $\rho_+\big(\mathcal{A}(u, w)\big) \le 1$ in $\overline{\Omega}$ is equivalent to $\varrho\big(A(u, w)\big) \le 1$ where

$$A(u, w) := \big(\nabla u, e(w)\big).$$

This way the problem $(\mathscr{P})$ will be recast as a particular case of $(\mathcal{P})$ after all.

Let us begin by studying the properties of the set:

$$\mathrm{C} := \left\{(\theta, \xi) \in \mathbb{R}^d \times \mathcal{S}^{d\times d} \,:\, \rho_+\left(\frac{1}{2}\,\theta \otimes \theta + \xi\right) \le 1\right\}. \tag{5.21}$$

A function $\bar{g} : \mathbb{R}^d \times \mathcal{S}^{d\times d} \to \mathbb{R}_+$ is defined and computed below:

$$\bar{g}(\theta, \xi) := \sup_{H\in\mathcal{K}_1} g\big(H, (\theta, \xi)\big) = \sup_{H\in\mathcal{K}_1} j_+\big(H, \tfrac{1}{2}\,\theta \otimes \theta + \xi\big) = \frac{1}{p}\left(\rho_+\left(\tfrac{1}{2}\,\theta \otimes \theta + \xi\right)\right)^p.$$

We thus recognize C as a sublevel set:

$$\mathrm{C} = \Big\{(\theta, \xi) \in \mathbb{R}^d \times \mathcal{S}^{d\times d} \,:\, \bar{g}(\theta, \xi) \le 1/p\Big\}. \tag{5.22}$$

The function $\bar{g}$ is a point-wise supremum of the family of functions $g(H, \cdot\,)$ that are convex continuous, cf. Lemma 5.3. Thus, by Proposition 2.8, $\bar{g}$ is convex and lower semi-continuous therefore C is a closed convex set that contains the origin. The function $\bar{g}$ is also continuous since ρ_+ is (see Proposition 3.11) and as a result C contains a neighbourhood of the origin. To C we assign its gauge

$$\varrho(\theta, \xi) := \inf_{t>0}\left\{t \,:\, \frac{1}{t}\,(\theta, \xi) \in \mathrm{C}\right\}$$

being a convex positively one-homogeneous continuous function (which is not a norm). Its polar ϱ^0 coincides with the support function of C:

$$\varrho^0(q, \sigma) = \sup_{(\theta,\xi)\in\mathbb{R}^d\times\mathcal{S}^{d\times d}} \Big\{\langle\theta, q\rangle + \langle\xi, \sigma\rangle : (\theta, \xi) \in \mathrm{C}\Big\} = \chi^*_{\mathrm{C}}(q, \sigma). \tag{5.23}$$

where we acknowledged yet another characterization of C

$$\mathrm{C} = \Big\{(\theta, \xi) \in \mathbb{R}^d \times \mathcal{S}^{d\times d} \,:\, \varrho(\theta, \xi) \le 1\Big\} \tag{5.24}$$

(see [140, Corollary 15.1.2] for a general relation between gauges and support functions).

Remark 5.11. It is worth stressing that, despite the resemblance between (5.21) and (5.24), for arbitrary $(\theta, \xi) \in \mathbb{R}^d \times \mathcal{S}^{d\times d}$ the value $\varrho(\theta, \xi)$ is not equal to $\rho_+\big(\frac{1}{2}\,\theta \otimes \theta + \xi\big)$. The argumentation is simple: clearly the mapping $(\theta, \xi) \mapsto \rho_+\big(\frac{1}{2}\,\theta \otimes \theta + \xi\big)$ is not positively one-homogeneous.

Lemma 5.12. *The following statements hold true:*

(i) The set C *is a closed (unbounded) convex subset of* $\mathbb{R}^d \times \mathcal{S}^{d\times d}$ *that contains a neighbourhood of the origin. Its support function reads*

$$\chi^*_{\mathrm{C}}(q,\sigma) = \varrho^0(q,\sigma) = \rho^0_+(\sigma) + \tfrac{1}{2}\langle \sigma^{-1}q, q\rangle \qquad (5.25)$$
$$= \begin{cases} \rho^0(\sigma) + \tfrac{1}{2}\langle \sigma\hat\theta, \hat\theta\rangle & \text{if } \sigma \in \mathcal{S}^{d\times d}_+ \text{ and } q = \sigma\hat\theta \text{ for } \hat\theta \in \mathbb{R}^d, \\ \infty & \text{if } \sigma \notin \mathcal{S}^{d\times d}_+ \text{ or if } q \notin \mathrm{Im}(\sigma). \end{cases}$$

(ii) Assume $(\theta,\xi) \in \mathbb{R}^d \times \mathcal{S}^{d\times d}$ *such that* $\varrho(\theta,\xi) \le 1$ *(equivalently* $(\theta,\xi) \in \mathrm{C}$ *or* $\rho_+(\frac{1}{2}\,\theta\otimes\theta + \xi) \le 1$*). Then, for any* $(q,\sigma) \in \mathbb{R}^d \times \mathcal{S}^{d\times d}$ *there holds the following characterization of the extremality relation:*

$$\langle\theta, q\rangle + \langle\xi,\sigma\rangle = \varrho^0(q,\sigma) \qquad \Leftrightarrow \qquad \begin{cases} \sigma \in \mathcal{S}^{d\times d}_+, \\ \left\langle \frac{1}{2}\,\theta\otimes\theta + \xi, \sigma\right\rangle = \rho^0(\sigma), \\ q = \sigma\theta. \end{cases}$$

Proof. The properties of C has been already established by exploiting (5.22). The proof of the rest of the statements is similar to the one of Lemma 5.3. We compute

$$\begin{aligned} \varrho^0(q,\sigma) &= \sup_{\theta\in\mathbb{R}^d}\sup_{\xi\in\mathcal{S}^{d\times d}} \Big\{\langle\theta,q\rangle + \langle\xi,\sigma\rangle : \rho_+\big(\tfrac{1}{2}\,\theta\otimes\theta + \xi\big) \le 1\Big\} \\ &= \sup_{\theta\in\mathbb{R}^d}\sup_{\tilde\xi\in\mathcal{S}^{d\times d}} \Big\{\langle\theta,q\rangle + \Big\langle\tilde\xi - \tfrac{1}{2}\,\theta\otimes\theta, \sigma\Big\rangle : \rho_+\big(\tilde\xi\,\big) \le 1\Big\} \\ &= \sup_{\theta\in\mathbb{R}^d} \Big\{\langle\theta,q\rangle - \tfrac{1}{2}\langle\sigma\theta,\theta\rangle\Big\} + \sup_{\tilde\xi\in\mathcal{S}^{d\times d}} \Big\{\langle\tilde\xi,\sigma\rangle : \rho_+\big(\tilde\xi\,\big) \le 1\Big\} = \tfrac{1}{2}\langle\sigma^{-1}q,q\rangle + \rho^0_+(\sigma) \end{aligned}$$

where we performed a substitution $\tilde\xi = \frac{1}{2}\,\theta\otimes\theta + \xi$. Under the prerequisite of the claim (ii) we clearly have $\rho_+(\tilde\xi) \le 1$. Then the assertion (ii) follows from the last line above while in addition we acknowledge the fact that $\rho^0_+(\sigma) < \infty$ if and only if $\sigma \in \mathcal{S}^{d\times d}_+$ and $\rho^0_+(\sigma) = \rho^0(\sigma)$ in this case. □

As announced the problem $(\mathscr{P})$ can now be rewritten by utilizing representation (5.24) of the set C:

$$\sup\mathscr{P} = \sup\Big\{\mathscr{F}(u,w) :\ (u,w) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1}),\ \ \varrho\big(A(u,w)\big) \le 1 \ \text{ in } \overline{\Omega}\Big\} \qquad (5.26)$$

where the linear functional and the linear operator read $\mathscr{F}(u,w) = \int u\,df$ and $A(u,w) = \big(\nabla u, e(w)\big)$ respectively. According to assertion (i) of Lemma 5.12 the function ϱ satisfies all the properties required for $\rho : V \to \mathbb{R}_+$ in Section 4.2 if we take $V = \mathbb{R}^d \times \mathcal{S}^{d\times d}$. The problem above is therefore a particular case of problem $(\mathcal{P})$ and automatically we have its dual $(\mathcal{P}^*)$ at our disposal, both are given in Section 4.2. The problem $(\mathcal{P}^*)$, where the unknown is a measure $m = (\vartheta,\varsigma) \in \mathcal{M}(\overline{\Omega};\mathbb{R}^d \times \mathcal{S}^{d\times d})$ (or in fact a pair of measures), reads:

$$\inf\Big\{\int \varrho^0(\vartheta,\varsigma)\ :\ (\vartheta,\varsigma) \in \mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d}),\ A^*(\vartheta,\varsigma) = \mathscr{F} \text{ in } \mathcal{D}'(\Omega;\mathbb{R}^{d+1})\Big\}. \qquad (5.27)$$

It is easy to check that the distributional equation $A^*(\vartheta,\varsigma) = \mathscr{F} = (f,0)$ in $\mathcal{D}'(\Omega;\mathbb{R}\times\mathbb{R}^d)$ reduces to a pair of equations $-\mathrm{div}\,\vartheta = f,\ \ -\mathrm{Div}\,\varsigma = 0$ in Ω. Since ϱ^0 is the support function

of the closed convex set C, by Corollary 2.24 the integral functional $(\vartheta,\varsigma)\mapsto\int\varrho^0(\vartheta,\varsigma)$ is convex weakly-* lower semi-continuous and it is the support function of the set of selections of C, i.e.

$$\int\varrho^0(\vartheta,\varsigma)=\sup_{(\phi,\psi)\in C(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d})}\left\{\int\langle\phi,\vartheta\rangle+\int\langle\psi,\varsigma\rangle:\big(\phi(x),\psi(x)\big)\in\mathrm{C}\ \ \forall\,x\in\overline{\Omega}\right\}. \quad (5.28)$$

For a pair $(\vartheta,\varsigma)\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d})$ let $\mu\in\mathcal{M}_+(\overline{\Omega})$ be any positive measure such that $\vartheta\ll\mu$ and $\varsigma\ll\mu$. Then, using the Goffman-Serrin convention and formula (5.25), for $(q,\sigma)\in L^1_\mu(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d})$ such that $(q\mu,\sigma\mu)=(\vartheta,\varsigma)$ there holds

$$\int\varrho^0(\vartheta,\varsigma)=\int\varrho^0(q,\sigma)\,d\mu=\int\Big(\rho^0_+(\sigma)+\tfrac12\langle\sigma^{-1}q,q\rangle\Big)d\mu. \quad (5.29)$$

Although the integral $\int\varrho^0(\vartheta,\varsigma)$ is well defined for any pair of measures $(\vartheta,\varsigma)\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d})$, the very particular form of the polar ϱ^0 entails a set of criteria under which the integral is finite:

Proposition 5.13. *In order that $\int\varrho^0(\vartheta,\varsigma)<\infty$, it is necessary that the following conditions hold true:*

(i) $\varsigma\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)$, *namely* $\varsigma\succeq0$;
(ii) $\vartheta\ll\rho^0(\varsigma)\in\mathcal{M}_+(\overline{\Omega})$, *i.e. for* $\mu:=\rho^0(\varsigma)$ *we have* $(\vartheta,\varsigma)=(q\mu,\sigma\mu)$ *for* $q\in L^1_\mu(\overline{\Omega};\mathbb{R}^d)$ *and* $\sigma\in L^\infty_\mu(\overline{\Omega};\mathcal{S}^{d\times d}_+)$ *with* $\rho^0(\sigma)=1\ \ \mu$*-a.e.*
(iii) for μ,q,σ *as above it holds that* $q\in\mathrm{Im}(\sigma)\ \ \mu$*-a.e.*

If the conditions above are met then there exists a μ-measurable function $\hat\theta:\overline{\Omega}\to\mathbb{R}^d$ such that $q=\sigma\hat\theta$ and the formulas below are valid:

$$\int\varrho^0(\vartheta,\varsigma)=\int\Big(\rho^0_+(\sigma)+\tfrac12\langle\sigma^{-1}q,q\rangle\Big)d\mu=\int\Big(\rho^0(\sigma)+\tfrac12\langle\sigma\hat\theta,\hat\theta\rangle\Big)d\mu$$
$$=\int\rho^0(\varsigma)+\int\tfrac12\,\langle\hat\theta\otimes\hat\theta,\varsigma\rangle.$$

Proof. By definition of ρ^0_+ the integral $\int\rho^0_+(\sigma)d\mu$ is finite only when $\sigma\in\mathcal{S}^{d\times d}_+\ \ \mu$-a.e., therefore the necessary condition $\varsigma\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)$ for finiteness of $\int\varrho^0(\vartheta,\varsigma)$ follows due to (5.29). Further we thus assume that $\varsigma\succeq0$. Next, for $\mu=\rho^0_+(\varsigma)=\rho^0(\varsigma)$ and $\sigma=\frac{d\varsigma}{d\mu}$, let $\vartheta=q\mu+q_sm_s$ be the Lebesgue decomposition of ϑ with respect to μ, in particular $\mu\perp m_s\in\mathcal{M}_+(\overline{\Omega})$. Then $\int\varrho^0(\vartheta,\varsigma)=\int\varrho^0(q,\sigma)\,d\mu+\int\varrho^0(q_s,0)dm_s$ and the condition for finiteness implies that $\int\varrho^0(q_s,0)dm_s=\int j^*_2(0,q_s)dm_s<\infty$ and consequently, by virtue of formula (5.10), that $q_s=0\ \ m_s$-a.e. Ultimately we must have $\vartheta=q\mu$ which we further assume. Then, if the set $B=\big\{q\notin\mathrm{Im}(\sigma)\big\}$ is of non-zero measure μ we have $\int_B\varrho^0(\vartheta,\varsigma)\ge\int_B\frac12\langle\sigma^{-1}q,q\rangle\,d\mu=\infty$ according to (5.10). Once $\mu(B)=0$, for μ-a.e. x we have $\frac12\big\langle(\sigma(x))^{-1}q(x),q(x)\big\rangle=\frac12\big\langle\sigma(x)\hat\theta(x),\hat\theta(x)\big\rangle$ for any $\hat\theta(x)$ satisfying $q(x)=\sigma(x)\,\hat\theta(x)$. Since point-wisely $\hat\theta$ solves the supremum problem in (5.10) the function $\hat\theta$ can be chosen as μ-measurable by virtue of [80, Chapter VII, Theorem 1.2]. Indeed, the integrand $(x,\theta)\mapsto\langle\theta,q\rangle-\frac12\big\langle\sigma(x)\theta,\theta\big\rangle$ enjoys the Carathéodory property. □

From Proposition 5.13 we can see that in the class of measures satisfying $\int\varrho^0(\vartheta,\varsigma)<\infty$ the measure ϑ – the absolute transverse force field in the membrane – relies on the structure

of measure ς – the absolute in-plane stress field in the membrane. This involved relation mirrors equation (iii) in (5.14) that governs the action of a membrane (see also equation (v) in Theorem 5.41, Section 5.4 below).

We may readily rewrite problem (5.27) as the dual of problem $(\mathscr{P})$:

$$\inf\left\{\int \varrho^0(\vartheta,\varsigma)\ :\ (\vartheta,\varsigma)\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d}_+),\ -\operatorname{div}\vartheta=f,\ \ \operatorname{Div}\varsigma=0\ \text{ in }\Omega\right\}\qquad(\mathscr{P}^*)$$

Note that the condition of positivity $\varsigma\succeq 0$ was explicitly written above, being justified by Proposition 5.13. Since the pair $(\mathscr{P}),(\mathscr{P}^*)$ is a particular case of the class of problems $(\mathcal{P}),(\mathcal{P}^*)$, the zero-gap result is now simply a specification of Proposition 4.3:

Proposition 5.14. *There holds a zero-gap equality*

$$\mathcal{Z}=\sup\mathscr{P}=\min\mathscr{P}^*$$

whereas the minimum is achieved.

Remark 5.15. The proof that $\mathcal{Z}<\infty$ will be given in Section 5.3.1.

5.2.4. Constructing solution of the (OEM) problem

We will now give the recipe for constructing an optimal elastic membrane $\check{\lambda}\in\mathcal{M}(\overline{\Omega};\mathcal{K})$ in the (OEM) problem based on a solution $(\vartheta,\varsigma)\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d}_+)$ of $(\mathscr{P}^*)$. As already stressed above the design problems (FMD) and (OEM) are different hence the route between solutions of $(\mathscr{P}^*)$ and (OEM) will diverge from the one put forth in Section 4.2, although there will be a lot of similarities. In particular the optimal material distribution $\check{\mu}=c(\check{\lambda})$ will not be proportional to measure $\varrho^0(\vartheta,\varsigma)$ (which could be expected when comparing problems $(\mathcal{P}^*)$ and $(\mathscr{P}^*)$) but to measure $\rho^0(\varsigma)$ instead.

Beforehand we require some additional insight into the features of solutions of $(\mathscr{P}^*)$:

Proposition 5.16. *The following facts hold true:*

(i) For $\varsigma\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)$ let $\mathcal{C}(\sigma)$ be the compliance functional defined in (5.1) *for the linear mebrane model. Then we have*

$$\int\rho^0(\varsigma)+\mathcal{C}(\varsigma)=\inf\left\{\int\varrho(\vartheta,\varsigma)\ :\ \vartheta\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d),\ -\operatorname{div}\vartheta=f\ \text{ in }\Omega\right\},\qquad(5.30)$$

whereas both sides of the equality may be infinite.

(ii) Let $(\hat{\vartheta},\hat{\varsigma})\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d}_+)$ be a solution to $(\mathscr{P}^*)$. Then the equi-repartition of energy rule holds true:

$$\frac{\mathcal{Z}}{2}=\int\rho^0(\hat{\varsigma})=\int\tfrac{1}{2}\langle\hat{\sigma}^{-1}\hat{q},\hat{q}\rangle\,d\hat{\mu}$$

where $\hat{\mu}=\rho^0(\hat{\varsigma})$, $\hat{\sigma}=\frac{d\hat{\varsigma}}{d\hat{\mu}}$ and $\hat{q}=\frac{d\hat{\vartheta}}{d\hat{\mu}}$.

Proof. Let us set $\mu=\rho^0(\varsigma)$ and $\sigma=\frac{d\varsigma}{d\mu}$. Owing to Proposition 5.13 the minimum on the right hand side of (5.30) may be restricted to measures of the form $\vartheta=q\mu$. The equality

$$\mathcal{C}(\varsigma)=\inf\left\{\int j_2^*(\sigma,q)\,d\mu\ :\ q\in L^2_\mu(\overline{\Omega};\mathbb{R}^d),\ -\operatorname{div}(q\mu)=f\text{ in }\Omega\right\}$$

is the statement of Proposition 4.4 if we set: $A := \nabla$, $\Upsilon_0 := \partial\Omega$, $\mathscr{F} := f$, $\lambda := \varsigma$, $\mathscr{H} := \sigma$, $\sigma := q$, $j : \mathcal{K} \times V \to \mathbb{R}_+$ as $j_2 : \mathcal{S}^{d\times d}_+ \times \mathbb{R}^d \to \mathbb{R}_+$. Assertion (i) follows since for $\varsigma \succeq 0$ we have $\int \varrho^0(q\mu, \sigma\mu) = \int \rho^0(\sigma)d\mu + \int j_2^*(\sigma, q)d\mu$.

Due to the claim (i) optimality of $(\hat\vartheta, \hat\varsigma)$ furnishes $\mathcal{Z} = \int \rho^0(\hat\varsigma) + \mathcal{C}(\hat\varsigma) = \min\big\{\int \rho^0(\varsigma) + \mathcal{C}(\varsigma) : \varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+),\ \mathrm{Div}\,\varsigma = 0 \text{ in } \Omega\big\}$. Then equality $\int \rho^0(\hat\varsigma) = \mathcal{C}(\hat\varsigma) = \mathcal{Z}/2$ can be proved in the identical way as for the statement (ii) of Proposition 5.1. □

The construction to be given relies on the anisotropy design tools developed in Chapter 3. For the reader's convenience we recall the notion that will be the most essential below: the set of Hooke tensors H that are optimal for a stress tensor $\sigma \in \mathcal{S}^{d\times d}_+$ equals

$$\underline{\mathcal{K}}_1(\sigma) = \arg\min\Big\{j^*(H, \sigma) : H \in \mathcal{K},\ c(H) \le 1\Big\},$$

while $H \in \underline{\mathcal{K}}_1(\sigma)$ if and only if $\frac{1}{p'}\big(\rho^0(\sigma)\big)^{p'} = j^*(H, \sigma)$ amongst $H \in \mathcal{K}$ with $c(H) \le 1$. One may notice that our membrane model assumes the constitutive model induced by the energy potential j_+ rather than j. However, if $\sigma \in \mathcal{S}^{d\times d}_+$ then $j^*_+(H, \sigma) = j^*(H, \sigma)$ and $\rho^0_+(\sigma) = \rho^0(\sigma)$, therefore $\underline{\mathcal{K}}_{1,+}(\sigma) = \underline{\mathcal{K}}_1(\sigma)$ in this case (see Proposition 3.26 for more details).

Theorem 5.17. *Let us consider a quadruple $(\hat\mu, \hat\sigma, \hat q, \hat{\mathscr{H}})$ such that $\hat\mu \in \mathcal{M}_+(\overline{\Omega})$, $\hat\sigma \in L^1_{\hat\mu}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$, $\hat q \in L^1_{\hat\mu}(\overline{\Omega}; \mathbb{R}^d)$ and $\hat{\mathscr{H}} \in L^\infty_{\hat\mu}(\overline{\Omega}; \mathcal{K})$. We define a quadruple $(\check\mu, \check\sigma, \check q, \check{\mathscr{H}})$ as below:*

$$\check\mu = \frac{2C_0}{\mathcal{Z}}\,\hat\mu, \qquad \check\sigma = \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{2p-2}{2p-1}}\hat\sigma, \qquad \check q = \frac{\mathcal{Z}}{2C_0}\,\hat q, \qquad \check{\mathscr{H}} = \hat{\mathscr{H}}. \tag{5.31}$$

Then, the two conditions are equivalent:

- *(i) the pair $(\hat\vartheta, \hat\varsigma) = \big(\hat q\hat\mu, \hat\sigma\hat\mu\big)$ solves problem $(\mathscr{P}^*)$ with $\rho^0(\hat\sigma) = 1$ $\hat\mu$-a.e. whilst $\hat{\mathscr{H}}(x) \in \underline{\mathcal{K}}_1\big(\hat\sigma(x)\big)$ for μ-a.e. x;*
- *(ii) the measure $\check\lambda = \check{\mathscr{H}}\check\mu$ solves the* (OEM) *problem with $c(\check{\mathscr{H}}) = 1$ $\check\mu$-a.e. whilst $(\check q, \check\sigma)$ solves the minimization problem in* (5.13) *for $\mu = \check\mu$ and $\mathscr{H} = \check{\mathscr{H}}$.*

Proof. First we assume that (i) holds true. Since $\check{\mathscr{H}} = \hat{\mathscr{H}} \in \underline{\mathcal{K}}_1\big(\hat\sigma(x)\big)$ we automatically have $c(\check{\mathscr{H}}) = 1$ $\hat\mu$-a.e. (note that $\hat\sigma \neq 0$ $\hat\mu$-a.e.). Admissibility of $\check\lambda = \check{\mathscr{H}}\check\mu$ can be then showed by exploiting the equi-repartition rule in the claim (ii) of Proposition 5.16:

$$\int c(\check\lambda) = \int c(\check{\mathscr{H}})\,d\check\mu = \int d\check\mu = \frac{2C_0}{\mathcal{Z}}\int d\hat\mu = \frac{2C_0}{\mathcal{Z}}\int \rho^0(\hat\varsigma) = C_0.$$

We shall now show that the pair $(\check q, \check\sigma)$ is admissible for the problem (5.13) with decomposition $\check\lambda = \check{\mathscr{H}}\check\mu$:

$$-\mathrm{div}\big(\check q\check\mu\big) = \frac{\mathcal{Z}}{2C_0}\,\frac{2C_0}{\mathcal{Z}}\Big(-\mathrm{div}\big(\hat q\hat\mu\big)\Big) = -\mathrm{div}\big(\hat q\hat\mu\big) = f \tag{5.32}$$

$$\mathrm{Div}(\check\sigma\check\mu) = \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{2p-2}{2p-1}}\frac{2C_0}{\mathcal{Z}}\,\mathrm{Div}(\hat\sigma\hat\mu) = \left(\frac{2C_0}{\mathcal{Z}}\right)^{\frac{1}{2p-1}}\mathrm{Div}(\hat\sigma\hat\mu) = 0 \tag{5.33}$$

Readily, by employing Proposition 5.6 we may estimate:

$$\begin{aligned}\mathscr{C}_{\min} \leq \mathscr{C}(\check{\lambda}) &\leq \int j^*(\check{\mathscr{H}},\check{\sigma})\,d\check{\mu} + \int \tfrac{1}{2}\langle \check{\sigma}^{-1}\check{q},\check{q}\rangle\,d\check{\mu}\\ &= \int \frac{1}{p'}\big(\rho^0(\check{\sigma})\big)^{p'}\,d\check{\mu} + \int \tfrac{1}{2}\langle \check{\sigma}^{-1}\check{q},\check{q}\rangle\,d\check{\mu}\\ &= \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{2p-1}}\left(\frac{1}{p'}\int d\hat{\mu} + \int \tfrac{1}{2}\langle \hat{\sigma}^{-1}\hat{q},\hat{q}\rangle\,d\hat{\mu}\right)\\ &= \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{2p-1}}\left(\frac{1}{p'}\frac{\mathcal{Z}}{2} + \frac{\mathcal{Z}}{2}\right) = \frac{2p-1}{2p}\left(\frac{\mathcal{Z}^{2p}}{2C_0}\right)^{\frac{1}{2p-1}} = \mathscr{C}_{\min}\end{aligned}$$

where:

- to pass to the second line we use the fact that $\check{\mathscr{H}}(x) = \hat{\mathscr{H}}(x) \in \mathcal{K}_1\big(\hat{\sigma}(x)\big) = \mathcal{K}_1\big(\check{\sigma}(x)\big)$;
- to pass to the forth line we acknowledge equality $\int d\hat{\mu} = \int \rho^0(\hat{\varsigma})$ and the equi-repartition of energy rule in assertion (ii) of Proposition 5.16;
- the last equality is due to Lemma 5.9.

Ultimately, the chain of inequalities above is a chain of equalities. Thus we infer that $\check{\lambda}$ is optimal for (OEM) and moreover that $(\check{q},\check{\sigma})$ solves (5.13).

We move on to showing the converse result, we assume (ii). Then admissibility $(\hat{\vartheta},\hat{\varsigma}) = \big(\hat{q}\hat{\mu},\hat{\sigma}\hat{\mu}\big)$ in the problem $(\mathscr{P}^*)$ is again due to (5.32), (5.33). Departing from optimality of $\check{\lambda}$ for (OEM) and of $(\check{q},\check{\sigma})$ for (5.13) we give a chain of estimates:

$$\begin{aligned}\mathscr{C}_{\min} = \mathscr{C}(\check{\lambda}) &= \int j^*(\check{\mathscr{H}},\check{\sigma})\,d\check{\mu} + \int \tfrac{1}{2}\langle \check{\sigma}^{-1}\check{q},\check{q}\rangle\,d\check{\mu}\\ &\geq \int \frac{1}{p'}\big(\rho^0(\check{\sigma})\big)^{p'}\,d\check{\mu} + \int \tfrac{1}{2}\langle \check{\sigma}^{-1}\check{q},\check{q}\rangle\,d\check{\mu}\\ &\geq \frac{1}{C_0^{p'/p}}\frac{1}{p'}\left(\int \rho^0(\check{\sigma})\,d\check{\mu}\right)^{p'} + \int \tfrac{1}{2}\langle \check{\sigma}^{-1}\check{q},\check{q}\rangle\,d\check{\mu}\\ &= \frac{1}{2}\left(\frac{\mathcal{Z}^{2p}}{2C_0}\right)^{\frac{1}{2p-1}}\left(\frac{1}{p'}\left(\frac{\int \rho^0(\hat{\sigma})\,d\hat{\mu}}{\mathcal{Z}/2}\right)^{p'} + \frac{\int \tfrac{1}{2}\langle \hat{\sigma}^{-1}\hat{q},\hat{q}\rangle\,d\hat{\mu}}{\mathcal{Z}/2}\right)\\ &\geq \frac{1}{2}\left(\frac{\mathcal{Z}^{2p}}{2C_0}\right)^{\frac{1}{2p-1}}\left(\frac{1}{p'}+1\right) = \frac{2p-1}{2p}\left(\frac{\mathcal{Z}^{2p}}{2C_0}\right)^{\frac{1}{2p-1}} = \mathscr{C}_{\min}\end{aligned}$$

where:

- to pass to the second line we use Proposition 3.40;
- to pass to the third line we use Hölder inequality (cf. the proof of Theorem 4.6);
- to pass to the last line we acknowledge that $\int \rho^0(\hat{\sigma})\,d\hat{\mu} + \int \tfrac{1}{2}\langle \hat{\sigma}^{-1}\hat{q},\hat{q}\rangle\,d\hat{\mu} \geq \min \mathscr{P}^* = \mathcal{Z}$ and for any positive number t:

$$\min_{a,b\in\mathbb{R}_+}\left\{\frac{1}{p'}\left(\frac{a}{t}\right)^{p'} + \frac{b}{t} \,:\, a+b \geq 2t\right\} = \frac{1}{p'}+1 \tag{5.34}$$

  while the unique minimizer of this problem is: $a = b = t$ (note that $p' > 1$).

Once again above we have equalities everywhere, which implies a series of results. First, from the analysis of the auxiliary problem (5.34) we deduce that $\int \rho^0(\hat{\sigma})\,d\hat{\mu} = \int \tfrac{1}{2}\langle \hat{\sigma}^{-1}\hat{q},\hat{q}\rangle\,d\hat{\mu} =$

$\mathcal{Z}/2$ and optimality of $(\hat{\vartheta}, \hat{\varsigma})$ for $(\mathscr{P}^*)$ follows. The remaining steps are analogous to those taken while proving the implication (ii) $\Rightarrow$ (i) in Theorem 4.6. $\square$

Remark 5.18. Similarly as in the case of Lemma 5.9, the proof of the result above uses the fact that the in-plane load F is not applied to the membrane (see (5.6)) and one can deduce this by comparing (5.32) and (5.33). We see from the latter equation that $\mathrm{Div}(\check{\sigma}\check{\mu})$ is scaled with respect to $\mathrm{Div}(\hat{\sigma}\hat{\mu})$, which would require rescaling the load F when passing between (OEM) and $(\mathscr{P}^*)$ as well.

By using the same arguments as for proving Corollary 4.7 we arrive at the method of constructing solutions to (OEM):

Corollary 5.19. *For $(\vartheta, \varsigma) \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d \times \mathcal{S}^{d\times d}_+)$ being any solution of $(\mathscr{P}^*)$ the measure $\mu = \rho^0(\varsigma)$ is Radon and we may define $\sigma = \frac{d\varsigma}{d\mu} \in L^\infty_\mu(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$.*

Let $\mathscr{H} \in L^\infty_\mu(\overline{\Omega}; \mathcal{K})$ be chosen as any μ-measurable selection of the multifunction $\overline{\Omega} \ni x \mapsto \underline{\mathcal{K}}_1\big(\sigma(x)\big)$, which always exists. Then, the constitutive field

$$\lambda = \frac{2C_0}{\mathcal{Z}}\, \mathscr{H}\mu$$

solves the (OEM) *problem.*

5.2.5. Optimal elastic membranes made of canvas-like material

So far in Section 5.2 the framework of optimal design of elastic membranes allowed an arbitrary choice of the triple $\mathcal{K}, j, c$ satisfying assumptions (A1)-(A4) and ($\mathrm{A_{ellip}}$) from Chapter 3. Now we shall concentrate on a particular choice of the triple that in Section 3.3.2 was referred to as the setting of Canvas-like Material Design (CanMD), cf. Example 3.28:

$$\mathcal{K} = \mathcal{K}_{fib}, \qquad j_+(H,\xi) \ \text{ for } \ j(H,\xi) = \tfrac{1}{2}\langle H\xi, \xi\rangle, \qquad c(H) = \mathrm{Tr}\, H. \tag{5.35}$$

where $\mathcal{K}_{fib} = \mathrm{co}\big(\big\{a\,\tau\otimes\tau\otimes\tau\otimes\tau \,:\, a \geq 0,\ \tau \in S^{d-1}\big\}\big)$. The assumption ($\mathrm{A_{ellip}}$) holds for the pair $\mathcal{K}_{fib}, j$, see Example 3.18.

This anisotropy setting is of paramount importance in the context of this chapter as well as the forthcoming one. First of all, in the author's opinion, this setting fits best the mechanical framework of membrane design. Membranes sewn as canvas are composed of large number of material fibres. If a fibre of cross-sectional area a is laid along a simple curve $\mathrm{C} \subset \overline{\Omega}$ then it is natural to assume that for a strain field $\epsilon \in C(\overline{\Omega}; \mathcal{S}^{d\times d})$ this fibre accumulates the elastic energy $\int j_+(\lambda^C, \epsilon)$ where $\lambda^C = a\,\tau\otimes\tau\otimes\tau\otimes\tau\,\mathcal{H}^1 \llcorner \mathrm{C}$ and τ is the vector tangent to C. Indeed, we see that point-wisely

$$j_+\big(\tau\otimes\tau\otimes\tau\otimes\tau, \epsilon\big) = \inf_{\zeta\in\mathcal{S}^{d\times d}_+} \frac{1}{2}\big(\langle \epsilon+\zeta, \tau\otimes\tau\rangle\big)^2 = \frac{1}{2}\Big(\big(\langle \epsilon, \tau\otimes\tau\rangle\big)_+\Big)^2, \tag{5.36}$$

namely only the tangential elongation contributes to the elastic energy, while the fibre is ideally compliant under contraction. Then, upon defining a space $\mathscr{X}(\overline{\Omega})$ of suitably regular curves (see [42, Section 3.2]), we can measure-theoretically simulate "sewing of

a canvas": for some $\alpha \in \mathcal{M}^b_+\big(\mathscr{X}(\overline{\Omega})\big)$ we construct a measure λ_α such that for every $\Psi \in C(\overline{\Omega}; \mathscr{L}(\mathcal{S}^{d\times d}))$ furnishes

$$\langle \Psi, \lambda_\alpha \rangle_{(C,\mathcal{M})} := \int_{\mathscr{X}(\overline{\Omega})} \left(\int_{\overline{\Omega}} \big\langle \Psi, \lambda^{\mathrm{C}} \big\rangle \right) \alpha(d\mathrm{C}). \tag{5.37}$$

Then, by assertion (ii) in Theorem 3.35, the constitutive field λ_α is an element of $\mathcal{M}(\overline{\Omega}; \mathcal{K}_{fib})$. Of course in the (OEM) problem in the CanMD setting we are searching the whole set $\big\{\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K}_{fib}) : \int c(\lambda) \le C_0\big\}$ therefore the solution $\check{\lambda}$ may not decompose to curves as in (5.37). Nonetheless, we shall learn that there are strong indications (both theoretical and numerical) that it is the case for a large class of loads and that we may restrict to curves being straight segments.

After Example 3.28 let us now recall the form of functions ρ_+, ρ^0_+ and the set $\underline{\mathcal{K}}_{1,+}(\sigma)$ in the CanMD setting. Note that in the case of FibMD, i.e. for j instead of j_+ in (5.35) we had

$$\rho(\xi) = \max_{i\in\{1,\dots,d\}} |\lambda_i(\xi)|, \qquad\qquad \rho^0(\sigma) = \sum_{i=1}^{d} |\lambda_i(\sigma)|,$$

while for CanMD there holds:

$$\rho_+(\xi) = \rho\big(\xi_+\big) = \max_{i\in\{1,\dots,d\}} \big(\lambda_i(\xi)\big)_+, \qquad\qquad \rho^0_+(\sigma) = \begin{cases} \operatorname{Tr}\sigma & \text{if } \sigma \in \mathcal{S}^{d\times d}_+, \\ \infty & \text{if } \sigma \notin \mathcal{S}^{d\times d}_+. \end{cases} \tag{5.38}$$

For any σ that is not positive semi-definite the set $\underline{\mathcal{K}}_{1,+}(\sigma)$ is trivially equal to the whole $\mathcal{K}_1$, while for $\sigma \in \mathcal{S}^{d\times d}_+$ the sets $\underline{\mathcal{K}}_1(\sigma)$ and $\underline{\mathcal{K}}_{1,+}(\sigma)$ are equal to each other and

$$\underline{\mathcal{K}}_{1,+}(\sigma) = \left\{ \sum_{i=1}^{m} \alpha_i\, \tau_i \otimes \tau_i \otimes \tau_i \otimes \tau_i \;:\; \tau_i \in \operatorname{Im}\sigma \cap S^{d-1},\ \sum_{i=1}^{m} \alpha_i = 1,\ \frac{\sigma}{\operatorname{Tr}\sigma} = \sum_{i=1}^{m} \alpha_i\, \tau_i \otimes \tau_i \right\}. \tag{5.39}$$

In Remark 3.29 it was stressed that $\underline{\mathcal{K}}_{1,+}(\sigma)$ contains plenty of Hooke tensors once $\operatorname{rank}(\sigma) > 1$: the optimal fibres may be laid along arbitrary direction τ_i in $\operatorname{Im}\sigma$.

We may readily point out the next feature that distinguishes the CanMD setting of the (OEM) problem: it is equivalent to the problem of designing Optimal Plasic Membrane (OPM). Indeed, by comparing (5.5) and (5.38) we deduce that

$$\rho_+\left(\frac{1}{2}\nabla u \otimes \nabla u + e(w)\right) \le 1 \qquad \Leftrightarrow \qquad \frac{1}{2}\nabla u \otimes \nabla u + e(w) \preceq \mathrm{I}.$$

Moreover, since $\rho^0(\sigma) = \operatorname{Tr}\sigma$ for $\sigma \in \mathcal{S}^{d\times d}_+$, from the claim (i) in Proposition 5.16 we can see that ς solves the problem (5.3) if and only if (ϑ, ς) solves $(\mathscr{P}^*)$ for some ϑ. As an effect $\widetilde{\mathcal{Z}} = \mathcal{Z}$ and equality holds true in (5.4).

Proposition 5.20. *Assume that $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K}) = \mathscr{H}\mu$ solves the* (OEM) *problem in the* CanMD *setting while $c(\mathscr{H}) = 1$ μ-a.e. Let $(q, \sigma) \in L^{(2p)'}_\mu(\overline{\Omega}; \mathbb{R}^d) \times L^{p'}_\mu(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ be a solution of the stress-based elasticity problem* (5.13)*. Then, the pre-stress field*

$$\varsigma = \frac{2\,\bar{s}\,V_0}{\mathcal{Z}} \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{2p-1}} \sigma\mu$$

solves the pre-stress design problem (OPM).

Proof. The quadruple $(\check{\mu}, \check{\sigma}, \check{q}, \check{\mathscr{H}}) = (\mu, \sigma, q, \mathscr{H})$ satisfies the condition (ii) in Theorem 5.17. Consequently, the pair $(\hat{\vartheta}, \hat{\varsigma}) = (\hat{q}\hat{\mu}, \hat{\sigma}\hat{\mu}) = \big(\check{q}\check{\mu}, (\frac{\mathcal{Z}}{2C_0})^{\frac{1}{2p-1}}\check{\sigma}\check{\mu}\big)$ solves $(\mathscr{P}^*)$ and therefore $\varsigma = \hat{\varsigma}$ solves (5.3). The claim now follows by assertion (iii) in Proposition 5.1. □

Considering the "if and only if" form of Theorem 5.17 it is easy to deduce the result that is converse to Proposition 5.20. All in all, we have succeeded to answer the doubts raised in the very first paragraph of Section 5.2. Indeed, we have shown that (up to a multiplicative constant) the pre-stress field ς, that is directly sought in the plastic design problem (OPM), can be recovered as an elastic response of membrane λ optimal for (OEM) to the out-of-plane load f. This matter will be further discussed, cf. Section 5.7.5.

5.3. Relaxation of the problem ($\mathscr{P}$)

In this section we address the issue of relaxing ($\mathscr{P}$) that is originally posed for smooth functions $(u, w) \in \mathcal{D}(\Omega; \mathbb{R}^{d+1})$. Owing to the reformulation (5.26) the problem ($\mathscr{P}$) is a particular case of the class of problems ($\mathcal{P}$) discussed in Chapter 4, therefore the relaxation process may run in accordance with Section 4.3. For the linear space $\mathcal{X} = \mathcal{D}(\Omega; \mathbb{R}^{d+1})$ we must thus choose a norm $|||\cdot|||$ such that

(B1) $\mathscr{F}(u, w) = \int u\, df = 0$ for any rigid body motion $(u, w) \in \mathcal{R} = \big\{(u, w) \in \mathcal{X} : A(u, w) = \big(\nabla u, e(w)\big) = (0, 0)\big\}$;

(B2) there exists $C > 0$ such that $|\mathscr{F}(u, w)| = \big|\int u\, df\big| \le C\, |||(u, w)|||$ for any $u \in \mathcal{U}$;

(B3) the set

$$\mathcal{X}_{1,\rho_+} := \left\{(u, w) \in \mathcal{D}(\Omega; \mathbb{R}^{d+1}) : \rho_+\left(\frac{1}{2}\,\nabla u \otimes \nabla u + e(w)\right) \le 1\right\}$$

is pre-compact in the Banach space $\overline{\mathcal{X}}$ being the completion of $\mathcal{X}$ in the norm $|||\cdot|||$.

It is easy to see that the space of rigid motions $\mathcal{R}$ defined above is equal to $\{(0, 0)\}$ due to the condition $\mathrm{spt}(u, w) \subset \Omega$, therefore (B1) is trivially satisfied. Having said that, the right choice of the norm $|||\cdot|||$ is not obvious at all. It is due to the fact that function j_+ does not satisfy the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$), see Proposition 3.23 and the comment above it. It is reflected in unboundedness of the set $\mathrm{C} = \big\{(\theta, \xi) : \rho_+\big(\frac{1}{2}\theta \otimes \theta + \xi\big) \le 1\big\}$; indeed, for any $\xi \in \mathcal{S}^{d\times d}_-$ and any $t \ge 0$ we have $(0, t\xi) \in \mathrm{C}$. This was not the case for frameworks #1 and #3 of the (FMD) problem put forth in Section 4.5 where condition ($\mathrm{A}_{\mathrm{ellip}}$) was enforced which enabled the choice of the uniform norm $|||\cdot||| = \|\cdot\|_{\infty,\Omega}$. In the case of the problem ($\mathscr{P}$) studied now the topology of uniform convergence is too strong to meet the condition (B3) (only for w as it will appear below).

The challenge of relaxing the problem ($\mathscr{P}$) is completely new. Below we will show that the correct norm is the following:

$$|||(u, w)||| := \|u\|_{\infty,\Omega} + \|w\|_{L^1(\Omega;\mathbb{R}^d)}. \tag{5.40}$$

We can see that with this particular choice validating condition (B2) is straightforward, which relies on the fact that the in-plane load F is zero, cf. (5.6). In what follows we will prove that the compactness condition (B3) holds true as well.

5.3.1. The compactness result for $(\mathscr{P})$

Instead of showing the compactness of the set $\mathcal{X}_{1,\rho_+}$ for general ρ_+ we shall concentrate only on the set

$$\mathcal{X}_1 := \left\{(u,w) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1}) \,:\, \frac{1}{2}\,\nabla u \otimes \nabla u + e(w) \preceq \mathrm{I} \ \text{ in } \overline{\Omega}\right\}$$

being the specification of $\mathcal{X}_{1,\rho_+}$ for ρ_+ given in (5.38), namely for the CanMD setting.

Proposition 5.21. *Assume that the pair $\mathcal{K}, j$ satisfies the ellipticity condition* ($\mathrm{A}_{\mathrm{ellip}}$). *Then there exists a constant $C > 0$ such that*

$$\frac{1}{C}\,\mathcal{X}_1 \subset \mathcal{X}_{1,\rho_+} \subset C\,\mathcal{X}_1.$$

Proof. Thanks to ($\mathrm{A}_{\mathrm{ellip}}$) the function ρ^0 constructed in (5.17) satisfies $\frac{1}{C_1}|\sigma| \le \ \rho^0(\sigma) \le \frac{1}{C_2}|\sigma|$ for $C_1, C_2 > 0$, see Proposition 3.13. It follows that all functions ρ^0 for $\mathcal{K}, j$ satisfying ($\mathrm{A}_{\mathrm{ellip}}$) are equivalent (in the sense similar to equivalence of norms). Owing to (5.16) we have

$$\rho_+(\xi) = \sup_{\sigma\in\mathcal{S}^{d\times d}} \left\{\langle \xi,\sigma\rangle : \rho^0_+(\sigma) \le 1\right\} = \sup_{\sigma\in\mathcal{S}^{d\times d}_+} \left\{\langle \xi,\sigma\rangle : \rho^0(\sigma) \le 1\right\}$$

and therefore ρ_+ in the class of pairs $\mathcal{K}, j$ meeting ($\mathrm{A}_{\mathrm{ellip}}$) are equivalent as well. The pair $\mathcal{K}, j$ specified in (5.35) is in this class, which furnishes the claim. □

The set $\mathcal{X}_1$ is exceptional comparing to other sets $\mathcal{X}_{1,\rho_+}$ owing to the following two-point characterization that is reminiscent of the one known for Michell problem (cf. Theorem 4.55):

Lemma 5.22. *For any smooth functions $(u,w) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1})$ (extended by zero to whole $\mathbb{R}^d$) the following conditions are equivalent:*

(i) the point-wise condition holds true

$$\frac{1}{2}\,\nabla u \otimes \nabla u + e(w) \preceq \mathrm{I} \qquad \textit{point-wise in } \mathbb{R}^d; \tag{5.41}$$

(ii) the two-point condition holds true

$$\frac{1}{2}\,|u(y)-u(x)|^2 + \langle w(y)-w(x), y-x\rangle \le |y-x|^2 \qquad \forall\,(x,y) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{5.42}$$

In other words $(u,w) \in \mathcal{X}_1$ if and only if the two-point condition (ii) is true.

Proof. First let us assume that the condition (i) is true. We choose a pair of distinct points $x, y \in \mathbb{R}^d$ and define $\tau^{x,y} = \frac{y-x}{|y-x|}$. We may write

$$\langle w(y)-w(x), \tau^{x,y}\rangle = \int_{[x,y]} \nabla_{\tau^{x,y}} \langle w, \tau^{x,y}\rangle \, d\mathcal{H}^1 = \int_{[x,y]} \langle e(w), \tau^{x,y}\otimes\tau^{x,y}\rangle \, d\mathcal{H}^1.$$

where for $\tau \in S^{d-1}$ by ∇_τ we understand the directional derivative. The estimate below holds due to Schwarz inequality

$$|u(y)-u(x)|^2 = \left|\int_{[x,y]} \nabla_{\tau^{x,y}} u \, d\mathcal{H}^1\right|^2 \le |y-x| \int_{[x,y]} \langle \nabla u\otimes\nabla u, \tau^{x,y}\otimes\tau^{x,y}\rangle \, d\mathcal{H}^1.$$

From the above we obtain

$$\begin{aligned} &\tfrac{1}{2}|u(y)-u(x)|^2+\langle w(y)-w(x),y-x\rangle \\ &\le |y-x|\int_{[x,y]}\big\langle \tfrac{1}{2}\,\nabla u\otimes\nabla u+e(w),\tau^{x,y}\otimes\tau^{x,y}\big\rangle\,d\mathcal{H}^1\le |x-y|^2. \end{aligned} \tag{5.43}$$

Conversely assume that (ii) holds and let us choose a point $x\in\mathbb{R}^d$ and a direction $\tau\in S^{d-1}$. By taking $y=x+h\,\tau$ for any $h>0$, directly from the two-point condition (5.42) we obtain

$$\frac{1}{2}\left(\frac{u(x+h\,\tau)-u(x)}{h}\right)^2+\left\langle\frac{w(x+h\,\tau)-w(x)}{h},\tau\right\rangle\le 1$$

which after taking the limit when $h\to 0$ yields

$$\tfrac{1}{2}\big(\nabla_\tau u(x)\big)^2+\langle\nabla_\tau w(x),\tau\rangle\le 1$$

or alternatively

$$\tfrac{1}{2}\,\langle\nabla u(x)\otimes\nabla u(x),\tau\otimes\tau\rangle+\langle e(w)(x),\tau\otimes\tau\rangle\le 1$$

which due to arbitrariness of τ and x furnishes (i) and completes the proof. □

Remark 5.23. From the proof we can see that the two-point condition (5.42) can be equivalently checked for pairs $(x,y)\in\overline{\Omega}\times\overline{\Omega}$.

In the two-point condition (5.42) lies an information that will be invaluable in the analysis of the set $\mathcal{X}_1$ and its suitable closure. For a pair $(u,w)\in\mathcal{X}_1$ let us define a function

$$v:=\mathrm{id}-w\quad\in\quad C^\infty(\mathbb{R}^d;\mathbb{R}^d)$$

where $\mathrm{id}=\mathrm{id}(x)=x$ is the identity map. Then, for any pair $(x,y)\in\mathbb{R}^d\times\mathbb{R}^d$ we have:

$$\begin{aligned} \langle v(y)-v(x),y-x\rangle &=\big\langle\big(y-x\big)-\big(w(y)-w(x)\big),y-x\big\rangle \\ &=|y-x|^2-\langle w(y)-w(x),y-x\rangle\ge\tfrac{1}{2}\,|u(y)-u(x)|^2\ge 0 \end{aligned} \tag{5.44}$$

In accordance with Section 2.4 the multifunction defined as $\mathbf{v}(x):=\{v(x)\}$ is a univalued maximal monotone map of full domain such that $\mathbf{v}=\mathrm{id}$ on $\mathbb{R}^d\setminus\overline{\Omega}$, namely

$$\mathbf{v}(x)=\{v(x)\}=\{x\}\qquad\forall\,x\in\mathbb{R}^d\setminus\overline{\Omega}.$$

This property of function v unlocks the results and tools developed for maximal monotone maps in [3], which are summarized in Section 2.4. In this subsection we shall use only one of those results (see the proof of lemma below), while we will return to exploit them more thoroughly in Section 5.3.2.

The estimates below will pave a way to relative compactness of the set $\mathcal{X}_1$ with respect to the norm (5.40):

Lemma 5.24. *Let D be the diameter of Ω. Then, for every $(u,w)\in\mathcal{X}_1$ it holds that:*

(i) $\|u\|_{\Omega,\infty}\le\frac{1}{\sqrt{2}}\,D,\ \ \|w\|_{\Omega,\infty}\le D$;
(ii) $|u(y)-u(x)|\le\sqrt{2D}\,|y-x|^{\frac{1}{2}}$ *for every* $(x,y)\in\overline{\Omega}\times\overline{\Omega}$;
(iii) $\int_\Omega|\mathrm{D}w|+\int_\Omega|\nabla u|^2\,d\mathcal{L}^d\le d\,|\Omega|+c_d\,D^d$ *(with c_d depending on d only).*

As a result $\mathcal{X}_1$ is a bounded convex subset of $\big(C^{0,\frac{1}{2}}(\Omega)\cap W^{1,2}(\Omega)\big)\times(W^{1,1}\cap L^\infty)(\Omega;\mathbb{R}^d)$.

Proof. Let us take an element $(u,w) \in \mathcal{X}_1$. By Lemma 5.22 the two point condition (5.42) holds for all $(x,y) \in \overline{\Omega} \times \overline{\Omega}$. In particular, we have $\langle w(y) - w(x), y - x\rangle \leq |x-y|^2$ while $w(y) = 0$ for all $y \in \partial\Omega$. Fix an arbitrary $x \in \Omega$. Then, for every every $\tau \in S^{d-1}$, there exists a unique real $t \in (0,D)$ such that $x' = x + t\tau \in \partial\Omega$. Thus we get $-\langle w(x), \tau\rangle \leq t \leq D$ and $|w|(x) \leq D$ as τ is arbitrary. This proves the second inequality in (i). Furthermore, if $x, y \in \Omega$ are distinct and $\tau^{x,y} = \frac{y-x}{|y-x|}$, there exists $s,t > 0$ such that $x' = x - s\tau^{x,y}$, $y' = y + t\tau^{x,y}$ belong to $\partial\Omega$ and $s+t+|y-x| = |y'-x'| \leq D$. We deduce from the above that $-\langle w(y) - w(x), \tau^{x,y}\rangle \leq s + t \leq D - |y-x|$, thus $-\langle w(y) - w(x), y - x\rangle \leq D|x-y| - |y-x|^2$. Eventually the assertion (ii) follows by again exploiting (5.42):

$$\frac{1}{2}\,|u(y) - u(x)|^2 \;\leq\; |x-y|^2 - \langle w(y) - w(x), y - x\rangle \;\leq\; D\,|x-y|.$$

As a result, the first inequality in (i) holds since $u = 0$ on $\partial\Omega$ and $\sup_{x\in\Omega} d(x,\partial\Omega) \leq D/2$.

Assertion (iii) is a consequence of applying statement (i) of Theorem 2.45 for $\mathscr{U} = \mathbb{R}^d$, $\mathscr{V} = \Omega$ and for the univalued maximal monotone map $\mathbf{v}(x) = \{v(x)\} = \{x - w(x)\}$. It states that there exists a constant $C_d > 0$ depending on the dimension only such that

$$\int_\Omega |\mathrm{D}v| \leq C_d D^{d-1}\,\mathrm{osc}(v,\Omega) \qquad \Big(\mathrm{osc}(v,\Omega) := \sup\big\{|v(x) - v(y)|,\ (x,y) \in \Omega^2\big\}\Big)$$

where $|\mathrm{D}v|$ is the variation measure for $v \in BV_{\mathrm{loc}}(\mathbb{R}^d;\mathbb{R}^d)$ (we assume it is computed with respect to ℓ^1 norm on $\mathbb{R}^{d\times d}$). By (i) the oscillation of v in Ω is not larger than $2D$ therefore $\int_\Omega |\mathrm{D}v| \leq 2C_d\,D^d$ and $\int_\Omega |\mathrm{D}w| \leq \int_\Omega |\mathrm{D}(\mathrm{id} - v)| \leq d\,|\Omega| + \int_\Omega |\mathrm{D}v|$. On the other hand, (5.41) implies that $\frac{1}{2}\,\nabla u \otimes \nabla u \preceq \mathrm{I} - e(w) = e(\mathrm{id} - w) = e(v)$, hence $\frac{1}{2}\,|\nabla u|^2 \leq \mathrm{Div}\, v$ by taking the traces. All in all, we deduce the estimate (iii) with $c_d = 6C_d$. □

The compactness result can be readily given:

Theorem 5.25. *Assume that the boundary $\partial\Omega$ is Lipschitz continuous. Then the set $\mathcal{X}_1$ is a relatively compact subset of $C(\overline{\Omega}) \times L^1(\Omega;\mathbb{R}^d)$.*

Let $\overline{\mathcal{X}}_1$ be the compact set that is the closure of $\mathcal{X}_1$ in the topology of $C(\overline{\Omega}) \times L^1(\Omega;\mathbb{R}^d)$. Then we have

$$\overline{\mathcal{X}}_1 \subset \Big((C^{0,\frac{1}{2}} \cap W^{1,2})(\Omega;\mathbb{R})\Big) \times \Big((BV \cap L^\infty)(\Omega;\mathbb{R}^d)\Big) \tag{5.45}$$

Moreover, if $(\mathrm{A}_{\mathrm{ellip}})$ *is satisfied for a pair $\mathcal{K}, j$, then the same holds true for the set $\mathcal{X}_{1,\rho_+}$ and its closure $\overline{\mathcal{X}}_{1,\rho_+}$ where ρ_+ is computed in* (5.15).

Proof. By Lemma 5.24 the set $\{u : (u,w) \in \mathcal{X}_1\}$ is bounded in $C^{0,\frac{1}{2}}(\Omega)$ hence its relative compactness in $C(\overline{\Omega})$ follows by Arzelà-Ascoli theorem. The additional boundedness in the reflexive space $W^{1,2}(\Omega)$ guarantees that the cluster points of the set (with respect to uniform convergence on Ω) lie in $C^{0,\frac{1}{2}}(\Omega) \cap W^{1,2}(\Omega)$. Next, thanks to Lipschitz continuity of $\partial\Omega$, the set $\{w : (u,w) \in \mathcal{X}_1\}$ being bounded in $W^{1,1}(\Omega)$ is pre-compact in $L^1(\Omega;\mathbb{R}^d)$ while its closure is contained in $BV(\Omega;\mathbb{R}^d)$, see [82, Section 5.2, Theorem 4]. In addition the upper bound on $\|w\|_{L^\infty(\Omega)} \leq D$ is preserved under closure in $L^1(\Omega;\mathbb{R}^d)$, hence (5.45). The "moreover part" of the statement follows directly by Proposition 5.21. □

We have proved that all the conditions (B1)-(B3) are satisfied for the norm (5.40). We are in a position to pose the relaxed variant of the problem ($\mathscr{P}$):

$$\max\left\{\int u\,df\ :\ (u,w)\in\overline{\mathcal{X}}_{1,\rho_+}\right\} \tag{$\overline{\mathscr{P}}$}$$

while the existence result is a particular case of Theorem 4.11:

Corollary 5.26. *Both the problems* ($\overline{\mathscr{P}}$) *and* ($\mathscr{P}^*$) *have solutions while*

$$\mathcal{Z}=\max\overline{\mathscr{P}}=\min\mathscr{P}^*<\infty.$$

As a result there exists a solution $\check\lambda$ *of* (OEM) *problem for which* $\mathscr{C}_{\min}=\mathscr{C}(\check\lambda)<\infty$.

In line with Section 4.3 we can now relate solution $(\hat u,\hat w)$ of the relaxed problem ($\overline{\mathscr{P}}$) to solution $(\check u,\check w)$ of the variational problem (5.9) defining compliance $\mathscr{C}(\lambda)$ for an optimally designed membrane $\lambda=\check\lambda$. We first define a *relaxed solution* of the problem (5.9):

Definition 5.27. Let $\check\lambda$ be a solution of the (OEM) problem. The functions $(\check u,\check w)\in C(\overline\Omega)\times L^1(\Omega;\mathbb{R}^d)$ shall be termed a relaxed solution of the maximization problem (5.9) if $(\check u,\check w)$ is a limit in $C(\overline\Omega)\times L^1(\Omega;\mathbb{R}^d)$ of a maximizing sequence (u_n,w_n) for problem (5.9) with $\lambda=\check\lambda$ and $\{(u_n,w_n)\}$ satisfies the upper bound (5.19).

In contrast to Proposition 4.13 the link between solutions $(\hat u,\hat w)$ and $(\check u,\check w)$ involves different scaling for u and w:

Proposition 5.28. *Let* $\check\lambda$ *be a solution of the* (OEM) *problem. For a function* $(\hat u,\hat w)\in\overline{\mathcal{X}}_1$ *we define*

$$\check u=\left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{2p-1}}\hat u,\qquad \check w=\left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{2}{2p-1}}\hat w$$

Then $(\hat u,\hat w)$ *is a solution of problem* ($\overline{\mathscr{P}}$) *if and only if* $(\check u,\check w)$ *is a relaxed solution of problem* (5.9) *for* $\lambda=\check\lambda$.

Proof. For $(\hat u,\hat w)$ solving ($\overline{\mathscr{P}}$) we have a maximizing sequence $(\hat u_n,\hat w_n)$ such that $\rho_+\big(\mathcal{A}(\hat u_n,\hat w_n)\big)\le 1$ and $\hat u_n\to u$ uniformly on $\overline\Omega$. We set $\check u_n=(\mathcal{Z}/(2C_0))^{\frac{1}{2p-1}}\hat u_n$ and $\check w_n=(\mathcal{Z}/(2C_0))^{\frac{2}{2p-1}}\hat w_n$ to discover that $\rho_+\big(\mathcal{A}(\check u_n,\check w_n)\big)\le(\mathcal{Z}/(2C_0))^{\frac{2}{2p-1}}$ and

$$\liminf_{n\to\infty}\left\{\int\check u_n\,df-\int j_+\big(\check\lambda,\mathcal{A}(\check u_n,\check w_n)\big)\right\}\ge\liminf_{n\to\infty}\left\{\int\check u_n\,df-\int\frac{1}{p}\Big(\rho_+\big(\mathcal{A}(\check u_n,\check w_n)\big)\Big)^p c(\check\lambda)\right\}$$

$$\ge\liminf_{n\to\infty}\left\{\int\check u_n\,df-\int\frac{1}{p}\left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{2p}{2p-1}}d\check\mu\right\}=\left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{2p-1}}\mathcal{Z}-\frac{C_0}{p}\left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{2p}{2p-1}}=\mathscr{C}(\check\lambda)$$

and therefore $(\check u_n,\check w_n)$ is a maximizing sequence for (5.9).

The prove of the converse implication is almost identical as in the case of Proposition 4.13 except for the energy repartition result: since the functional $u\mapsto j_+\big(\check\lambda,\mathcal{A}(u,w)\big)$ is positive $2p$-homogeneous we have $\int\check u\,df=(2p)'\,\mathscr{C}(\check\lambda)=\frac{2p}{2p-1}\mathscr{C}(\check\lambda)=(\mathcal{Z}^{2p}/(2C_0))^{\frac{1}{2p-1}}$ and therefore $\int\hat u\,df=\mathcal{Z}$. □

5.3.2. Characterization of the set $\overline{\mathcal{X}}_1$ through maximal monotone maps

Let us turn our attention to the monotonicity property of functions $v = \mathrm{id} - w$ discovered for $(u, w) \in \mathcal{X}_1$, see (5.44). More precisely we have shown that the univalued multifunction $\mathbf{v} = \mathbf{v}(x) = \{v(x)\}$ belongs to the set

$$\mathbf{V}_\Omega := \Big\{ \mathbf{v} : \mathbb{R}^d \to 2^{\mathbb{R}^d} \ : \ \mathbf{v} \text{ maximal monotone, } \mathbf{v} = \mathrm{id} \quad \text{in } \mathbb{R}^d \setminus \overline{\Omega} \Big\},$$

whereas a multifunction is monotone if it satisfies

$$\langle y' - x', y - x \rangle \geq 0 \qquad \begin{array}{l} \forall\, (x, y) \in \mathbb{R}^d \times \mathbb{R}^d, \\ \forall\, x' \in \mathbf{v}(x),\ \forall\, y' \in \mathbf{v}(y); \end{array}$$

for the notion of maximality and properties of maximal monotone maps $\mathbf{v}$ see Section 2.4. Let us stress that by id we will denote the identity function (then $\mathrm{id}(x) = x$) or the univalued identity multifunction (then $\mathrm{id}(x) = \{x\}$) depending on the context.

We shall exploit the foregoing link to maximal monotone maps to provide a characterization of the set $\overline{\mathcal{X}}_1$, i.e. the set that we search in the problem $(\overline{\mathscr{P}})$ in the setting of Canvas-like Material Design (CanMD). The exceptional mathematical features of $\overline{\mathcal{X}}_1$ that will be herein established constitute yet another reason of why we concentrate on this very particular case of the Optimal Elastic Membrane problem.

From Section 2.4 we find that the set of maximal monotone maps is strongly related to a set of vector valued functions of locally bounded variation, i.e. a subset of $BV_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d)$. We recall that the distributional derivative of such a function *is locally a matrix valued Radon measure*, which we write $\mathrm{D}v \in \mathcal{M}(\mathbb{R}^d; \mathbb{R}^{d\times d})$ understanding that $\mathrm{D}v \llcorner K$ is a bounded Radon measure for any compact $K \subset \mathbb{R}^d$. Its symmetrical part will be denoted by

$$\varepsilon(v) := \frac{1}{2}\Big(\mathrm{D}v + (\mathrm{D}v)^\top\Big) \quad \in \quad \mathcal{M}(\mathbb{R}^d; \mathcal{S}^{d\times d}).$$

We introduce a set of BV functions:

$$\mathcal{V}_\Omega := \Big\{ v \in BV_{\mathrm{loc}} \cap L^\infty_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d) \ : \ \varepsilon(v) \succeq 0, \quad v = \mathrm{id} \ \text{ a.e. in } \mathbb{R}^d \setminus \overline{\Omega} \Big\}.$$

In the condition $\varepsilon(v) \succeq 0$, which may be rewritten as $\varepsilon(v) \in \mathcal{M}(\mathbb{R}^d; \mathcal{S}^{d\times d}_+)$, one can find the monotonicity property of v enforced a.e. The rigorous relation between the two sets $\mathbf{V}_\Omega$ and $\mathcal{V}_\Omega$ is the following:

Proposition 5.29. *Let D be the diameter of Ω. The following statements hold true:*

- *(i) Each map $\mathbf{v} \in \mathbf{V}_\Omega$ is of full domain, i.e. $\mathbf{v}(x) \neq \varnothing$ for each $x \in \mathbb{R}^d$, and $\mathbf{v}$ is upper semi-continuous as a multifunction.*
- *(ii) For every $\mathbf{v} \in \mathbf{V}_\Omega$ there exists a unique $v \in \mathcal{V}_\Omega$ such that $\mathbf{v}(x) = \{v(x)\}$ for a.e. $x \in \mathbb{R}^d$. Conversely, for every $v \in \mathcal{V}_\Omega$ there exists a unique $\mathbf{v} \in \mathbf{V}_\Omega$ such that $\mathbf{v}(x) = \{v(x)\}$ for a.e. $x \in \mathbb{R}^d$.*
- *(iii) If $\mathbf{v} \in \mathbf{V}_\Omega$ then $\mathbf{v}(x) \subset \overline{B}(x, D)$ for every $x \in \mathbb{R}^d$, while $x \in \mathbf{v}(x)$ if $x \in \partial\Omega$;*
- *(iv) There exists a constant C_d depending on the dimension d only such that*

$$\int_\Omega |\mathrm{D}v| \leq 2C_d D^d \qquad \text{for every } v \in \mathcal{V}_\Omega. \tag{5.46}$$

Proof. By the very definition of $\mathbf{V}_\Omega$ we have $\mathbf{v}(x) = x$ in $\mathbb{R}^d \setminus \overline{\Omega}$ therefore $\operatorname{dom}\mathbf{v} \subset \mathbb{R}^d \setminus \overline{\Omega}$. By virtue of the claim (iii) in Proposition 2.39 we have $\operatorname{int}\big(\operatorname{co}(\operatorname{dom}\mathbf{v})\big) \subset \operatorname{dom}\mathbf{v}$ therefore $\operatorname{dom}\mathbf{v} = \mathbb{R}^d$ by boundedness of Ω. Then the upper semi-continuity property on whole $\mathbb{R}^d$ is due to claim (iv) in Proposition 2.39. Assertion (i) follows. Taking (i) into consideration the statement (ii) follows from Theorem 2.45 for $\mathscr{U} = \mathbb{R}^d$. For given $\mathbf{v} \in \mathbf{V}_\Omega$ the fact that $x \in \mathbf{v}(x)$ for $x \in \overline{\Omega}$ is a consequence of upper semi-continuity of the multifunction v and the fact that $\mathbf{v}(x) = \{x\}$ outside $\overline{\Omega}$. To prove the first part of (iii) we define $\mathbf{w} = \mathrm{id} - \mathbf{v}$; then from monotonicity of $\mathbf{v}$ we infer that $\langle y' - x', y - x\rangle \leq |y - x|^2$ for any $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ and any $x' \in \mathbf{w}(x),\ y' \in \mathbf{w}(y)$ while $0 \in \mathbf{w}(x)$ if $x \in \partial\Omega$. By adapting the first step of the proof of Lemma 5.24 to multifunction $\mathbf{w}$ we deduce that $|x'| \leq D$ for any $x \in \overline{\Omega}$ and $x' \in \mathbf{w}(x)$, while clearly $\mathbf{w}(x) = \{0\}$ in $\mathbb{R}^d \setminus \overline{\Omega}$. In the end $\mathbf{w}(x) \subset \overline{B}(0, D)$ for each $x \in \mathbb{R}^d$. The statement (iii) follows since $\mathbf{v} = \mathrm{id} - \mathbf{w}$. Assertion (iv) is due to Theorem 2.45 used for $\mathscr{U} = \mathbb{R}^d$ and $\mathscr{V} = \Omega$. Indeed, the oscillation $\operatorname{osc}(\mathbf{v}, \Omega)$ (cf. Theorem 2.45 for definition) is not greater that $2D$ by assertion (iii) of this result. □

We shall now endow the sets $\mathbf{V}_\Omega$ and $\mathcal{V}_\Omega$ with metrics. For any pair $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}_\Omega$ we define

$$\mathbf{h}(\mathbf{v}_1, \mathbf{v}_2) := D_H\big(\mathcal{G}_{\mathbf{v}_1}, \mathcal{G}_{\mathbf{v}_2}\big).$$

where D_H is the Hausdorff distance (see (2.15) in Section 2.3) and $\mathcal{G}_{\mathbf{v}_i} = \big\{(x, x') : x \in \mathbb{R}^d, x' \in \mathbf{v}_i(x)\big\}$ is the graph of the multifunction $\mathbf{v}_i$. By assertion (ii) in Proposition 2.39 $\mathcal{G}_{\mathbf{v}_i}$ are closed therefore $\mathbf{h}$ is an extended metric on $\mathbf{V}_\Omega$ by Proposition 2.28. From assertion (iii) in Proposition 5.29 we deduce that $\mathbf{h}(\mathbf{v}_1, \mathbf{v}_2) < \infty$ therefore $(\mathbf{V}_\Omega, \mathbf{h})$ it is a classical metric space (the word *extended* is redundant). Next, for any pair $v_1, v_2 \in \mathcal{V}_\Omega$ we put

$$d_{L^1}(v_1, v_2) := \|v_1 - v_2\|_{L^1(\Omega;\mathbb{R}^d)} = \int_\Omega |v_1 - v_2|\, d\mathcal{L}^d.$$

which again is a metric, since $v_1 = v_2 = \mathrm{id}$ a.e in $\mathbb{R}^d \setminus \overline{\Omega}$. It is also finite since $v(x) \in \overline{B}(x, D)$ for a.e. $x \in \mathbb{R}^d$ by (iii) in Proposition 5.29. In view of statement (ii) in Proposition 5.29, we may consider the one-to-one mapping between metric spaces

$$\mathbf{i} : (\mathbf{V}_\Omega, \mathbf{h}) \ni \mathbf{v} \ \mapsto\ v \in (\mathcal{V}_\Omega, d_{L^1}).$$

Lemma 5.30. *The metric spaces* $(\mathbf{V}_\Omega, \mathbf{h})$ *and* $(\mathcal{V}_\Omega, d_{L^1})$ *are compact while* $\mathbf{i}$ *is a homeomorphism. Moreover, the metric* $\mathbf{h}$ *induces the graph convergence* (2.16)

Proof. Let $\mathbf{M}$ be the set of all maximal monotone maps on $\mathbb{R}^d$, while by $\emptyset$ denote the trivial multifunction whose graph is an empty set. Then, according to Theorem 2.41, the set $\mathbf{M} \cup \{\emptyset\}$ is compact with respect to an extended metric $\mathbf{d}$ that induces the graph convergence (2.16). Let us now show that the set $\mathbf{V}_\Omega \subset \mathbf{M}$ is a compact subset of $\mathbf{M} \cup \{\emptyset\}$. We take a sequence $\{\mathbf{v}_n\}$ such that $\mathbf{v}_n \to \mathbf{v} \in \mathbf{M} \cup \{\emptyset\}$ in the sense of graph convergence. Since $\mathbf{v}_n = \mathrm{id}$ in $\mathbb{R}^d \setminus \Omega$ it is straightforward to check that $\mathbf{v} = \mathrm{id}$ therein as well, therefore $\mathbf{v} \neq \emptyset$ and $\mathbf{v} \in \mathbf{M}$ as a result. Ultimately $\mathbf{v} \in \mathbf{V}_\Omega$ by definition hence the closedness and thus compactness of $\mathbf{V}_\Omega$ with respect to metric $\mathbf{d}$. By virtue of the "moreover part" of Theorem 2.41 the topology in $\mathbf{V}_\Omega$ does not change if we replace $\mathbf{d}$ with $\mathbf{h}$, which proves compactness of the space $(\mathbf{V}_\Omega, \mathbf{h})$ and that its topology induces graph convergence.

Let us now prove that the one-to-one mapping $\mathbf{i}$ is continuous. Assume that $\mathbf{v}_n \to \mathbf{v}$ in $\mathbf{V}_\Omega$ in the sense of graph and consider a Lebesgue negligible Borel subset N such that $\mathbf{v}_n(x) = \{v_n(x)\}$ and $\mathbf{v}(x) = \{v(x)\}$ for all $x \notin N$ and for every $n \in \mathbb{N}_+$. Since $v_n(x) \in \overline{B}(x, D)$ the sequence $\{(x, v_n(x))\}$ admits a cluster point (x, x') and by property (i) of graph convergence (2.16) there must necessarily hold $x' = v(x)$. We thus infer that $v_n \to v$ a.e. in Ω and, by the dominated convergence theorem, the convergence $d(v_n, v) = \int_\Omega |v_n - v|\, dx \to 0$ follows. Since $\mathbf{h}$ is continuous and $(\mathbf{V}_\Omega, \mathbf{h})$ is compact we deduce that $\mathbf{h}$ is a homeomorphism and $(\mathcal{V}_\Omega, d_{L^1})$ is compact as well. □

The above results, established for monotone maps $\mathbf{v}$ and their BV counterparts v, can be now traced back to multifunction $\mathbf{w} = \mathrm{id} - \mathbf{v}$ and another BV function $w = \mathrm{id} - v$. We defines the sets

$$\mathbf{W}_\Omega := \mathrm{id} - \mathbf{V}_\Omega, \qquad \mathcal{W}_\Omega := \mathrm{id} - \mathcal{V}_\Omega$$

and we endow them with the metrics $\mathbf{h}$, d_{L^1} respectively. Then we may define the one to one mapping $\tilde{\mathbf{i}} : \mathbf{W}_\Omega \to \mathcal{W}_\Omega$ by putting $\tilde{\mathbf{i}}(\mathbf{w}) := \mathrm{id} - \mathbf{i}(\mathrm{id} - \mathbf{w})$. In what follows we will slightly abuse the notation by writing $\mathbf{i}$ in the place of $\tilde{\mathbf{i}}$.

Let us agree on convention regarding decomposition of derivatives of BV functions. For any $w \in BV_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ by $\mathrm{D}w = \mathrm{D}_a w + \mathrm{D}_s w \in \mathcal{M}(\mathbb{R}^d; \mathbb{R}^{d\times d})$ we shall understand the Lebesgue decomposition to the absolute continuous part $\mathrm{D}_a w \ll \mathcal{L}^d$ and the singular part $\mathrm{D}_s w \perp \mathcal{L}^d$. On top of that by $\nabla w \in L^1_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^{d\times d})$ we shall understand the density of $\mathrm{D}_a w$, namely $\mathrm{D}_a w = \nabla w\, \mathcal{L}^d$. Similarly for $\varepsilon(w) = \frac{1}{2}\big(\mathrm{D}w + (\mathrm{D}w)^\top\big) \in \mathcal{M}(\mathbb{R}^d; \mathcal{S}^{d\times d})$ we will understand notation $\varepsilon(w) = \varepsilon_a(w) + \varepsilon_s(w)$ and $\varepsilon(w) = e(w)\, \mathcal{L}^d$ for $e(w) \in L^1(\mathbb{R}^d; \mathcal{S}^{d\times d})$.

Proposition 5.31. *The sets $\mathbf{W}_\Omega, \mathcal{W}_\Omega$ enjoy the following properties:*

(i) The metric spaces $\big(\mathbf{W}_\Omega, \mathbf{h}\big)$ and $\big(\mathcal{W}_\Omega, d_{L^1}\big)$ are compact and $\mathbf{i} : \mathbf{W}_\Omega \to \mathcal{W}_\Omega$ is a homeomorphism.

(ii) A multifunction $\mathbf{w}$ is an element of $\mathbf{W}_\Omega$ if and only if $\mathbf{w}(x) = \{0\}$ for $x \in \mathbb{R}^d \setminus \overline{\Omega}$ and

$$\langle y' - x', y - x\rangle \le |y - x|^2 \qquad \begin{array}{l}\forall\, (x, y) \in \mathbb{R}^d \times \mathbb{R}^d,\\ \forall\, x' \in \mathbf{w}(x),\ \forall\, y' \in \mathbf{w}(y).\end{array} \tag{5.47}$$

Moreover: every $\mathbf{w} \in \mathbf{W}_\Omega$ is an upper semi-continuous multifunction; for any $x \in \mathbb{R}^d$ $\mathbf{w}(x)$ is a non-empty closed convex set contained in the closed ball $\overline{B}(0, D)$ where D is the diameter of Ω.

(iii) For any pair $(u, w) \in \overline{\mathcal{X}}_1$ the function w belongs to $\mathcal{W}_\Omega$.

(iv) Let $\mathbf{w} \in \mathbf{W}_\Omega$ and $w = \mathbf{i}(\mathbf{w}) \in \mathcal{W}_\Omega$. Then there exists a $\mathcal{L}^d$-negligible set $N \subset \overline{\Omega}$ such that on $\mathbb{R}^d \setminus N$ the map $\mathbf{w}$ is univalued and differentiable in the following sense: for every $x_0 \in \mathbb{R}^d \setminus N$ there exists $\nabla\mathbf{w}(x_0) \in \mathbb{R}^{d\times d}$ such that

$$\lim_{\substack{x\to x_0\\ x'\in\mathbf{w}(x)}} \frac{x' - \mathbf{w}(x_0) - \nabla\mathbf{w}(x_0)(x - x_0)}{|x - x_0|} = 0. \tag{5.48}$$

Moreover, N can be chosen so that $\nabla\mathbf{w} = \nabla w$ on $\mathbb{R}^d \setminus N$.

(v) Every function $w \in \mathcal{W}_\Omega$ belongs to $BV \cap L^\infty(\mathbb{R}^d; \mathbb{R}^d)$ while

$$e(w) \preceq \mathrm{I} \quad \text{a.e. in } \mathbb{R}^d, \qquad \varepsilon_s(w) \preceq 0 \quad \Big(\text{namely } \varepsilon_s(w) \in \mathcal{M}(\mathbb{R}^d; \mathcal{S}^{d\times d}_-)\Big).$$

Proof. The statement (i) is straightforward. The multifunction $\mathbf{w}$ belongs to $\mathbf{W}_\Omega$ if and only if $\mathbf{v} := \mathrm{id} - \mathbf{w}$ belongs to $\mathbf{V}_\Omega$. As a result the condition $\mathbf{w} = \{0\}$ outside $\overline{\Omega}$ in assertion (ii) is clear; moreover we see that $\mathbf{v}$ is monotone if and only if for any $x' \in \mathbf{w}(x)$, $y' \in \mathbf{w}(y)$

$$0 \leq \big\langle \big(\mathrm{id}(y) - y'\big) - \big(\mathrm{id}(x) - x'\big), y - x \big\rangle = \langle y - x, y - x\rangle - \langle y' - x', y - x\rangle \tag{5.49}$$

which establishes the first part of assertion (ii). Its "moreover part" is due to the fact that $\mathbf{w} = \mathrm{id} - \mathbf{v}$ for $\mathbf{v} \in \mathbf{V}_\Omega$ for which statements (i), (iii) in Proposition 5.29 hold true.

If $(u, w) \in \mathcal{X}_1$ then by Lemma 5.22 and assertion (ii) of this proposition the univalued multifunction $\mathbf{w} := \{w\}$ is an element of $\mathbf{W}_\Omega$ and therefore $w = \mathbf{i}^{-1}(\mathbf{w}) \in \mathcal{W}_\Omega$. The claim (iii) follows by compactness of $\mathcal{W}_\Omega$ in the L^1-topology.

The first part of the assertion (iv) follows directly from Theorem 2.43 if we acknowledge that $\mathbf{v} = \mathrm{id} - \mathbf{w} \in \mathbf{V}_\Omega$ (note that $\nabla \mathbf{w} = 0$ in $\mathbb{R}^d \setminus \overline{\Omega}$). The "moreover part" of (iv) is implied by the fact that a function $w \in BV_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{R}^d)$ is a.e. approximatively differentiable and the approximate derivative $\operatorname{ap} \nabla w$ a.e. coincides with ∇w (see [82, Chapter 6, Theorem 4]). Indeed, by comparing the definition of the approximate derivative and (5.48) it is clear that $\operatorname{ap} \nabla w(x) = \nabla \mathbf{w}(x)$ at x for which both the derivatives exist.

Ultimately we prove (v). By definition of $\mathcal{W}_\Omega$ we have $v = \mathrm{id} - w \in \mathcal{V}_\Omega$ therefore $\varepsilon_a(w) = \varepsilon_a(\mathrm{id} - v) = \varepsilon_a(\mathrm{id}) - \varepsilon_a(v)$ while $\varepsilon_s(w) = \varepsilon_s(\mathrm{id}) - \varepsilon_s(v)$. Since $\varepsilon_a(v)$, $\varepsilon_s(v)$ are mutually singular measures the condition $\varepsilon(v) \in \mathcal{M}(\mathbb{R}^d; \mathcal{S}^{d\times d}_+)$ implies that both $\varepsilon_a(v)$, $\varepsilon_s(v) \in \mathcal{M}(\mathbb{R}^d; \mathcal{S}^{d\times d}_+)$. On top of that we find that $\varepsilon_a(\mathrm{id}) = \mathrm{I}\,\mathcal{L}^d$ and $\varepsilon_s(\mathrm{id}) = 0$. Eventually $-\varepsilon_s(w) \in \mathcal{M}(\mathbb{R}^d; \mathcal{S}^{d\times d}_+)$ and $e(v) = \mathrm{I} - e(w) \in \mathcal{S}^{d\times d}_+$ a.e. in $\mathbb{R}^d$. Clearly $w \in L^\infty(\mathbb{R}^d)$ since by assertion (ii) the multifunction $\mathbf{w} = \mathbf{i}^{-1}(w)$ satisfies $\mathbf{w}(x) \subset \overline{B}(0, D)$ for every $x \in \mathbb{R}^d$. The proof of the statement (v) is complete. □

We note that functions $(u, w) \in \mathcal{X}_1$ are in fact defined on the whole $\mathbb{R}^d$ since $(u, w) \in \mathcal{D}(\Omega; \mathbb{R}^{d+1})$. Then, the elements $(u, w) \in \overline{\mathcal{X}}_1$ that are *a priori* defined on $\overline{\Omega}$ may be viewed as $u \in C_0(\mathbb{R}^d)$ and $w \in L^1(\mathbb{R}^d; \mathbb{R}^d)$ when extended by zero to whole $\mathbb{R}^d$ (note that $u = 0$ on $\partial\Omega$ as u is a limit of $u_h \in \mathcal{D}(\Omega)$ in the topology of uniform convergence). Ultimately the set $\overline{\mathcal{X}}_1$ may be equivalently defined as the closure of $\mathcal{X}_1$ in $C_0(\mathbb{R}^d) \times L^1(\mathbb{R}^d; \mathbb{R}^d)$.

We are now in a position to characterize the set $\overline{\mathcal{X}}_1$. The author believes that the same result holds for Ω with Lipschitz continuous boundary, however, in order not to dwell on technical issues related to the boundary we will restrain ourselves to domains that are star-shaped with respect to a ball (see Definition 4.30):

Theorem 5.32 (Characterization of the set $\overline{\mathcal{X}}_1$). *Assume a bounded domain Ω that is star-shaped with respect to a ball. Let us be given functions $u \in C_0(\mathbb{R}^d)$ and $w \in L^1(\mathbb{R}^d; \mathbb{R}^d)$ such that $(u, w) = (0, 0)$ in $\mathbb{R}^d \setminus \overline{\Omega}$. Then the following conditions are equivalent:*

(i) $(u, w) \in \overline{\mathcal{X}}_1$;

(ii) there exists a multifunction $\mathbf{w} \in \mathbf{W}_\Omega$ *such that* $w = \mathbf{i}(\mathbf{w})$ *and*

$$\frac{1}{2}\,|u(y) - u(x)|^2 + \langle y' - x', y - x\rangle \leq |y - x|^2 \qquad \begin{array}{l} \forall\,(x, y) \in \mathbb{R}^d \times \mathbb{R}^d, \\ \forall\, x' \in \mathbf{w}(x),\ \forall\, y' \in \mathbf{w}(y); \end{array} \tag{5.50}$$

(iii) $u \in W^{1,2}(\mathbb{R}^d)$ *and* $w \in BV \cap L^\infty(\mathbb{R}^d;\mathbb{R}^d)$ *whilst*

$$\varepsilon_s(w) \text{ is a negative semi-definite matrix valued Radon measure,} \tag{5.51}$$

$$\frac{1}{2}\nabla u \otimes \nabla u + e(w) \preceq \mathrm{I} \qquad \text{a.e. in } \mathbb{R}^d \tag{5.52}$$

where ∇u *is the weak gradient and* $e(w) = \frac{d\varepsilon_a(w)}{d\mathcal{L}^d}$.

Moreover, if any of the conditions above is met, the function u *is a.e. differentiable in the classical sense and this derivative of* u *coincides a.e. with the weak derivative* $\nabla u \in L^2(\mathbb{R}^d;\mathbb{R}^d)$.

Remark 5.33. The condition (ii) in Theorem 5.32 is equivalent to the following: there exists a maximal monotone map $\mathbf{v} \in \mathbf{V}_\Omega$ such that $w = \mathrm{id} - \mathbf{i}(\mathbf{v})$ and

$$\frac{1}{2}|u(y)-u(x)|^2 \le \langle y'-x', y-x\rangle \qquad \begin{array}{l}\forall\,(x,y) \in \mathbb{R}^d \times \mathbb{R}^d,\\ \forall\, x' \in \mathbf{v}(x),\ \forall\, y' \in \mathbf{v}(y),\end{array} \tag{5.53}$$

which can be deduced through a chain of inequalities similar to (5.49). Similarly, under transformation $v = \mathrm{id} - w$ condition (iii) translates to

$$\varepsilon_s(v) \text{ is a positive semi-definite matrix valued Radon measure,} \tag{5.54}$$

$$\frac{1}{2}\nabla u \otimes \nabla u \preceq e(v) \qquad \text{a.e. in } \mathbb{R}^d. \tag{5.55}$$

Remark 5.34. The two-point condition (5.53) can be equivalently tested for pairs of points (x,y) in $\overline{\Omega}\times\overline{\Omega}$ only. Indeed, if it is met on the set $\overline{\Omega}\times\overline{\Omega}$ then for $(x,y) \in (\mathbb{R}^d \setminus \overline{\Omega})^2$ the condition follows trivially, while for $(x,y) \in (\mathbb{R}^d \setminus \overline{\Omega}) \times \overline{\Omega}$ we take $\bar{x}$ as any element of the set $[x,y] \cap \partial\Omega$ to find that for any $y' \in \mathbf{v}(y)$ we have

$$\begin{aligned}\frac{1}{2}|u(y)-u(x)|^2 &= \frac{1}{2}|u(y)-u(\bar{x})|^2 \le \langle y'-\bar{x}, y-\bar{x}\rangle\\ &\le \langle y'-\bar{x}, y-\bar{x}\rangle + \langle \bar{x}-x, y-\bar{x}\rangle = \langle y'-x, y-\bar{x}\rangle \le \langle y'-x, y-x\rangle\end{aligned}$$

where we used the fact that $u(\bar{x}) = u(x) = 0$ and $\bar{x} \in \mathbf{v}(\bar{x})$, $\mathbf{v}(x) = \{x\}$. Thus the same can be asserted for the two-point condition (5.50), see also Remark 5.23 for the smooth variant.

First we shall separately prove the differentiation property for u:

Lemma 5.35. *For* $u \in C_0(\mathbb{R}^d)$ *assume that there exists a multifunction* $\mathbf{w} \in \mathbf{W}_\Omega$ *such that the two-point condition* (5.50) *holds true. Then* u *is differentiable in the classical sense a.e. in* $\mathbb{R}^d$.

Proof. For N being the null set from assertion (iv) in Proposition 5.31 let us define $A = \Omega \setminus N$, i.e. at every point $x_0 \in A$ there exists the derivative $\nabla\mathbf{w}(x_0)$ as in (5.48). Based on the latter condition we may easily infer that for each $x_0 \in A$ there exists a radius $R(x_0) > 0$ and a constant $C(x_0) > 0$ such that

$$|x_2' - x_1'| \le C(x_0)|x_2 - x_1| \qquad \forall\, x_i \in B\big(x_0, R(x_0)\big), \quad \forall\, x_i' \in \mathbf{w}(x_i) \quad \text{for } i \in \{1,2\},$$

which by combining with the two-point condition (5.50) yields

$$\begin{aligned}\tfrac{1}{2}\,|u(x_2)-u(x_1)|^2 &\le |x_2-x_1|^2-\langle x_2'-x_1',x_2-x_1\rangle\\ &\le \big(1+C(x_0)\big)|x_2-x_1|^2 \qquad\qquad \forall\, x_i\in B\big(x_0,R(x_0)\big)\end{aligned}$$

rendering the function u Lipschitz continuous in each open ball $B\big(x_0,R(x_0)\big)$ for each $x_0\in A$ with the Lipschitz constant $\sqrt{2+2\,C(x_0)}$. By Rademacher's theorem (see [82, Chapter 3, Theorem 2]) for each $x_0\in A$ there exists a null set $\hat{N}(x_0)$ such that u is classically differentiable in the set $B\big(x_0,R(x_0)\big)\backslash\hat{N}(x_0)$. We introduce a family of open balls $\mathcal{G}=\big\{B(x_0,r)\,:\,x_0\in A,\ r\le R(x_0)\big\}$ being an open cover of the set A. Thanks to Lindelöf's lemma (see [109, Chapter 1, Theorem 15]) there exists a countable subfamily $\mathcal{G}_0=\big\{B(x_j,r_j)\,:\,j\in\mathbb{N},\ x_j\in A,\ r_j\le R(x_j)\big\}\subset\mathcal{G}$ of balls such that $A\subset\bigcup_{B(x_j,r_j)\in\mathcal{G}_0}B(x_j,r_j)$. We define the set $\hat{A}=\bigcup_{B(x_j,r_j)\in\mathcal{G}}B(x_j,r_j)\backslash\hat{N}(x_j)$; clearly u is differentiable in the set $\hat{A}$ and at the same time

$$\begin{aligned}\mathcal{L}^d\big(\Omega\backslash\hat{A}\big)&\le\mathcal{L}^d\big(\Omega\backslash A\big)+\mathcal{L}^d\big(A\backslash\hat{A}\big)=\mathcal{L}^d\big(A\backslash\hat{A}\big)\\ &\le\mathcal{L}^d\Big(\bigcup_{B(x_j,r_j)\in\mathcal{G}}\hat{N}(x_j)\Big)\le\sum_{B(x_j,r_j)\in\mathcal{G}}\mathcal{L}^d\big(\hat{N}(x_j)\big)=0,\end{aligned}$$

where at the end we utilized the fact that the sum is countable and that $\hat{N}(x_j)$ is a null set for every x_j. We have thus showed that u is differentiable in the classical sense a.e. in Ω and therefore also in whole $\mathbb{R}^d$ since $u=0$ in $\mathbb{R}^d\backslash\Omega$. □

Proof of Theorem 5.32. Proof of implication (i) ⇒ (ii): Let us take a pair $(u,w)\in\mathcal{X}_1$, then by definition we have at our disposal a sequence of smooth functions $\big\{(u_n,w_n)\big\}\subset\mathcal{X}_1$ such that $(u_n,w_n)\to(u,w)$ in $C(\mathbb{R}^d)\times L^1(\mathbb{R}^d;\mathbb{R}^d)$. Let us denote a sequence of univalued multifunctions $\mathbf{w}_n=\mathbf{i}^{-1}(w_n)$, then by assertion (iii) of Proposition 5.31 we obtain that $\mathbf{w}_n\in\mathbf{W}_\Omega$ for any h and thus, thanks to assertion (i) of Proposition 5.31, there exists a subsequence (without relabelling) $\mathbf{w}_n$ such that $\mathbf{w}_n\to\mathbf{w}$ in $\mathbf{W}_\Omega$, i.e. in the sense of graph. Next, for a pair $x_1,x_2\in\mathbb{R}^d$ let us take any $x_1'\in\mathbf{w}(x_1)$ and $x_2'\in\mathbf{w}(x_2)$; since the sequence $\mathbf{w}_n$ converges to $\mathbf{w}$ as graphs we deduce that there exist sequences $x_{1,n},x_{2,n}\in\mathbb{R}^d$ such that

$$\big(x_{i,n},w_n(x_{i,n})\big)\to(x,x_i')\quad\text{in }\mathbb{R}^d\times\mathbb{R}^d\qquad\text{for } i\in\{1,2\},$$

where we have used the fact that $\mathbf{w}_n(x)=\{w_n(x)\}$ for any $x\in\mathbb{R}^d$ and h. Further, due to the uniform convergence of $u_n\to u$ we obtain that $|u_n(x_{i,n})-u(x_i)|\to 0$ for $i\in\{1,2\}$. By applying Lemma 5.22 to each pair $(u_n,w_n)\in\mathcal{X}_1$ we infer that

$$\begin{aligned}&|x_2-x_1|^2-\tfrac{1}{2}|u(x_2)-u(x_1)|^2-\langle x_2'-x_1',x_2-x_1\rangle\\ &=\lim_{h\to\infty}\left(|x_{2,n}-x_{1,n}|^2-\frac{1}{2}|u_n(x_{2,n})-u_n(x_{1,n})|^2-\langle w_n(x_{2,n})-w_n(x_{1,n}),x_{2,n}-x_{1,n}\rangle\right)\ge 0\end{aligned}$$

furnishing (5.50). It is thus left to show that $\mathbf{i}(\mathbf{w})=w$. Since $\mathbf{w}_n=\mathbf{i}^{-1}(w_n)$ and $\mathbf{w}_n\to\mathbf{w}$ in $\mathbf{W}_\Omega$ then w_n converges to $\mathbf{i}(\mathbf{w})$ in $\mathcal{W}_\Omega$, i.e. in $L^1(\mathbb{R}^d;\mathbb{R}^d)$, owing to $\mathbf{i}$ being a homeomorphism, cf. assertion (i) in Proposition 5.31. Hence $\mathbf{i}(\mathbf{w})$ must be nothing but w.

Proof of implication (ii) ⇒ (iii): For (u,w) satisfying the prerequisite of the theorem we assume that there exists $\mathbf{w}\in\mathbf{W}_\Omega$ such that (5.50) holds. The fact that w is an element of $BV\cap L^\infty(\mathbb{R}^d;\mathbb{R}^d)$ that satisfies (5.51) is a direct implication of statement (v)

in Proposition 5.31. We shall now prove that $u \in W^{1,2}(\mathbb{R}^d)$. In fact, since $u \in C_0(\Omega)$ it is enough to show that $u \in W^{1,2}(\Omega)$. We set $\mathbf{v} = \mathrm{id} - \mathbf{w} \in \mathbf{V}_\Omega$ (a maximal monotone map) and $v = \mathrm{id} - w = \mathbf{i}(\mathbf{v}) \in \mathcal{V}_\Omega$. Let $\omega \Subset \Omega$ and fix $h \in \mathbb{R}^d$ such that $|h| < \mathrm{dist}(\omega, \partial\Omega)$. As the set $\big\{x \in \omega : \mathbf{v}(x) = \{v(x)\},\ \mathbf{v}(x+h) = \{v(x+h)\}\big\}$ is of full Lebesgue measure in ω (see Theorem 2.42), the condition (5.50) implies that:

$$\frac{1}{2}\big(u(x+h) - u(x)\big)^2 \le \langle v(x+h) - v(x), h\rangle \qquad \text{for a.e. } x \in \omega, \tag{5.56}$$

cf. Remark 5.33. By integrating over ω we deduce that

$$\frac{1}{2}\int_\omega \big(u(x+h) - u(x)\big)^2\,dx \ \le\ |h|^2 \int_\omega \left|\frac{v(x+h) - v(x)}{|h|}\right| dx \ \le\ C\,|h|^2\ ,$$

where the constant C depends only on $\int_\Omega |\mathrm{D}v|$ (the variation of v as an element of $BV(\Omega;\mathbb{R}^d)$) and thus it is finite by virtue of Proposition 5.29. Since ω and h with $|h| < \mathrm{dist}(\omega, \partial\Omega)$ can be chosen arbitrarily, the latter uniform upper bound entails that $u \in W^{1,2}(\Omega)$ (see for instance [48, Proposition 9.3] and the subsequent Remark 6).

By combining Lemma 5.35 and assertion (iv) of Proposition 5.31 we deduce that there exists a set $A \subset \Omega$ of full Lebesgue measure in Ω such that for every $x \in A$: the function u is classically differentiable at x and the derivative $\nabla u(x)$ coincides with the weak gradient; $\mathbf{w}(x) = \{w(x)\}$ and the multifunction $\mathbf{w}$ is differentiable at x in the sense (5.48) and this derivative $\nabla\mathbf{w}(x)$ coincides with the density ∇w of $\mathrm{D}_a w$ at x. As a result for every $x \in A$ and any direction $\tau \in S^{d-1}$ there holds (recall that $e(w)(x) = \frac{1}{2}\big(\nabla w(x) + (\nabla w(x))^\top\big)$):

$$\begin{aligned}&\big\langle \tfrac{1}{2}\,\nabla u(x) \otimes \nabla u(x) + e(w)(x), \tau\otimes\tau\big\rangle = \tfrac{1}{2}\,|\langle \nabla u(x), \tau\rangle|^2 + \langle \nabla\mathbf{w}(x)\,\tau, \tau\rangle\\ &= \lim_{\substack{h\to 0_+\\ y'\in\mathbf{w}(x+h\,\tau)}} \frac{1}{2}\left(\frac{u(x+h\,\tau) - u(x)}{h}\right)^2 + \left\langle \frac{y' - w(x)}{h}, \tau\right\rangle \le 1\end{aligned}$$

where the latter inequality comes from the inequality (5.50) evaluated for $x, y = x + h\tau$ and divided on both sides by $h^2 = |y - x|$. The above chain of inequalities implies condition (5.52) in the set $\Omega \setminus A$ and therefore a.e. in Ω. Meanwhile, in the set $\mathbb{R}^d \setminus \overline{\Omega}$ inequality (5.52) trivially holds true as $(u, w) = (0, 0)$ therein. The proof of the second implication is now established.

Proof of implication (iii) $\Rightarrow$ (i). We take a function $u \in C(\mathbb{R}^d) \cap W^{1,2}(\mathbb{R}^d)$ and $w \in L^\infty(\mathbb{R}^d;\mathbb{R}^d) \cap BV(\mathbb{R}^d;\mathbb{R}^d)$ such that $u = 0$, $w = 0$ a.e. in $\mathbb{R}^d\setminus\overline{\Omega}$ and (5.51), (5.52) hold true. Our goal is to construct a sequence $(u_n, w_n) \in \mathcal{X}_1 \subset \mathcal{D}(\Omega;\mathbb{R}^{d+1})$ such that $u_n \to u$ uniformly in $\mathbb{R}^d$ and $w_n \to w$ in $L^1(\mathbb{R}^d;\mathbb{R}^d)$. We recall that condition (5.52) may be rewritten as: $\bar g\big(\nabla u, e(w)\big) \le 1$ a.e. where $\bar g : \mathbb{R}^d \times \mathcal{S}^{d\times d} \to \mathbb{R}_+$ is a convex continuous function satisfying

$$\bar g(\theta, \xi) = \rho_+\big(\tfrac{1}{2}\,\theta\otimes\theta + \xi\big) = \max\left\{\lambda_1\left(\tfrac{1}{2}\,\theta\otimes\theta + \xi\right), 0\right\}, \tag{5.57}$$

see (5.38). Without loss of generality we may assume that Ω is star shaped with respect to a ball centred at the origin. For $\delta > 0$ we define the dilations of functions u, w:

$$u_\delta(x) := \frac{1}{1+\delta}\,u\big((1+\delta)x\big), \qquad w_\delta(x) := \frac{1}{1+\delta}\,w\big((1+\delta)x\big).$$

It is clear that still $u_\delta \in C(\mathbb{R}^d) \cap W^{1,2}(\mathbb{R}^d)$ and $w_\delta \in L^\infty(\mathbb{R}^d)^d \cap BV(\mathbb{R}^d)^d$ while $u_\delta \to u$ uniformly on $\mathbb{R}^d$ and $w_\delta \to w$ in $L^1(\mathbb{R}^d;\mathbb{R}^d)$ when $\delta \to 0$. It is also easy to check that u_δ, w_δ satisfy (5.51) and (5.52) that is:

$$\varepsilon_s(w_\delta) \in \mathcal{M}(\mathbb{R}^d;\mathcal{S}^{d\times d}_-) \qquad \text{and} \qquad \bar{g}\big(\nabla u_\delta, e(w_\delta)\big) \leq 1 \quad \text{a.e. in } \mathbb{R}^d.$$

Finally $(u,w) = 0$ in $\mathbb{R}^d \setminus \Omega_\delta$ where the dilated domain $\Omega_\delta = \frac{1}{1+\delta}\,\Omega$ satisfies $D_H(\partial\Omega, \Omega_\delta) =: \varepsilon_\delta > 0$ (cf. the comment below Definition 4.30).

For $0 < \varepsilon < \varepsilon_\delta$ let us take the standard mollification kernel η_ε whose support is contained in $B(0,\varepsilon)$. The mollified functions $u_{\delta,\varepsilon} := \eta_\varepsilon * u_\delta$, $w_{\delta,\varepsilon} := \eta_\varepsilon * w_\delta$ are then compactly supported in Ω, i.e. $(u_{\delta,\varepsilon}, w_{\delta,\varepsilon}) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1})$ and, owing to the classical results (see [82]), $u_{\delta,\varepsilon} \to u_\delta$ uniformly in $\mathbb{R}^d$ and $w_{\delta,\varepsilon} \to w_\delta$ in $L^1(\mathbb{R}^d;\mathbb{R}^d)$ when $\varepsilon \to 0$. Finally, in order to prove that $(u_{\delta,\varepsilon}, w_{\delta,\varepsilon}) \in \mathcal{X}_1$ we must show that $\bar{g}\big(\nabla u_{\delta,\varepsilon}, e(w_{\delta,\varepsilon})\big) \leq 1$ in Ω. Since $u \in W^{1,2}(\mathbb{R}^d)$ and $w \in BV(\mathbb{R}^d;\mathbb{R}^d)$ it is well establishes that $\nabla u_{\delta,\varepsilon} = \eta_\varepsilon * \nabla u_\delta$ and $\varepsilon(w_{\delta,\varepsilon}) = \eta_\varepsilon * \varepsilon(w_\delta) = \eta_\varepsilon * \varepsilon_a(w_\delta) + \eta_\varepsilon * \varepsilon_s(w_\delta) = \big(\eta_\varepsilon * e(w_\delta)\big)\,\mathcal{L}^d + \eta_\varepsilon * \varepsilon_s(w_\delta)$ where $e(w_\delta) = \frac{d\varepsilon_a(w_\delta)}{d\mathcal{L}^d}$. Since $\varepsilon_s(w_\delta) \in \mathcal{M}(\mathbb{R}^d;\mathcal{S}^{d\times d}_-)$ it is straightforward that $\eta_\varepsilon * \varepsilon_s(w_\delta) \in \mathcal{S}^{d\times d}_-$ point-wise in $\mathbb{R}^d$. The following chain of inequalities hold for every $x \in \mathbb{R}^d$:

$$\begin{aligned}
&\bar{g}\big(\nabla u_{\delta,\varepsilon}(x), e(w_{\delta,\varepsilon})(x)\big) = \bar{g}\bigg(\big(\eta_\varepsilon * \nabla u_\delta\big)(x)\ ,\ \big(\eta_\varepsilon * e(w_\delta)\big)(x) + \big(\eta_\varepsilon * \varepsilon_s(w_\delta)\big)(x)\bigg)\\
&\leq \bar{g}\bigg(\big(\eta_\varepsilon * \nabla u_\delta\big)(x)\ ,\ \big(\eta_\varepsilon * e(w_\delta)\big)(x)\bigg) = \bar{g}\bigg(\int_{\mathbb{R}^d}\Big(\nabla u_\delta(y), e(w_\delta)(y)\Big)\,\eta_\varepsilon(x-y)\,\mathcal{L}^d(dy)\bigg)\\
&\leq \int_{\mathbb{R}^d} \bar{g}\Big(\nabla u_\delta(y), e(w_\delta)(y)\Big)\,\eta_\varepsilon(x-y)\,\mathcal{L}^d(dy) \leq 1
\end{aligned}$$

where:

- to pass to the second line we used the fact that for $\theta \in \mathbb{R}^d$, $\xi_1 \in \mathcal{S}^{d\times d}$ and $\xi_2 \in \mathcal{S}^{d\times d}_-$ there holds $\bar{g}(\theta, \xi_1 + \xi_2) \leq \bar{g}(\theta, \xi_1)$; indeed, $\lambda_1\big(\frac{1}{2}\theta\otimes\theta + \xi_1 + \xi_2\big) \leq \lambda_1\big(\frac{1}{2}\theta\otimes\theta + \xi_1\big)$ due to the negative semi-definiteness of ξ_2;
- to pass to the third line Jensen's inequality was employed for the convex function $\bar{g} : \mathbb{R}^d \times \mathcal{S}^{d\times d} \to \mathbb{R}_+$.

We have thus showed that $(u_{\delta,\varepsilon}, w_{\delta,\varepsilon}) \in \mathcal{X}_1$. Owing to the aforementioned convergences in δ and ε the sequence $(u_n, w_n) := (u_{\delta_n,\varepsilon_n}, w_{\delta_n,\varepsilon_n})$ converging to (u,w) in $C_0(\mathbb{R}^d) \times L^1(\mathbb{R}^d;\mathbb{R}^d)$ may be found by using the diagonalization argument. The proof of the last implication, and consequently of the whole theorem, is complete. □

Remark 5.36. The two-point characterization (ii) of the set $\overline{\mathcal{X}}_1$ allows to transfer its properties to lower dimensional affine subsets of Ω. Let us consider an affine k-dimensional ($k \leq d$) subspace A of $\mathbb{R}^d$ such that $\Omega_A := A \cap \Omega$ is non-empty. In addition let V be the k-dimensional linear subspace of $\mathbb{R}^d$ such that $A = x_0 + V$ for some $x_0 \in A$, while for $P_V \in \mathcal{S}^{d\times d}$ we take the orthogonal projection operator onto V. Then for any pair $(u,w) \in \overline{\mathcal{X}}_1$ we put $u_A := u|_A$ and $\mathbf{w}_A := \big(P_V(\mathbf{i}^{-1}w)\big)\big|_A$. Clearly we have $u_A \in C_0(\Omega_A)$

and $\mathbf{w}_A : A \to 2^V$ is an upper semi-continuous convex valued multifunction. From the characterization (ii) in Theorem 5.32 we may easily infer that

$$\frac{1}{2}\,|u_A(y) - u_A(x)|^2 + \langle y' - x', y - x\rangle \le |y - x|^2 \qquad \begin{array}{l}\forall\,(x,y) \in A \times A,\\ \forall\, x' \in \mathbf{w}_A(x),\ \forall\, y' \in \mathbf{w}_A(y)\end{array} \tag{5.58}$$

where we used the fact that $P_V(y-x) = y - x$. In particular we see that $\mathbf{v}_A := P_V\,\mathrm{id} - \mathbf{w}_A$ is monotone on A. We can define the respective notions $\overline{\mathcal{X}}_1(\Omega_A)$, $\mathbf{W}_{\Omega_A}$, $\mathcal{W}_{\Omega_A}$ and the homeomorphism $\mathbf{i}_A$ to see that $(u_A, w_A) \in \overline{\mathcal{X}}_1(\Omega_A)$ for $w_A = \mathbf{i}_A(\mathbf{w}_A) \in L^1(\Omega_A; V)$ (the Lebesgue space is taken with respect to the $\mathcal{H}^k$ measure). Through condition (iii) in Theorem 5.32 we infer that

$$u_A \in W^{1,2}(\Omega_A), \qquad w_A \in BV(\Omega_A; V).$$

while

$$\frac{1}{2}\,\nabla_A u_A \otimes \nabla_A u_A + e_A(w_A) \preceq P_V$$

where $\nabla_A u_A$ is a classical tangential weak gradient and $e_A(w)$ is defined upon the distributional tangential derivative $\mathrm{D}_A w_A \in \mathcal{M}(A; V \overset{s}{\otimes} V)$, cf. the convention put forth above Proposition 5.31.

We will now make use of the dimension reduction property discussed in Remark 5.36 by taking A to be a straight line in $\mathbb{R}^d$. Let us fix a pair of distinct points $(a, b) \in \overline{\Omega} \times \overline{\Omega}$ and $\tau^{a,b} = \frac{b-a}{|b-a|}$. For a pair $(u, w) \in \overline{\mathcal{X}}_1$ let us define a univariate real function $u_{a,b} : \mathbb{R} \to \mathbb{R}$ and a multifunction $\mathbf{v}_{a,b} : \mathbb{R} \to 2^{\mathbb{R}}$

$$u_{a,b}(t) := u\big(a + t\,\tau^{a,b}\big), \qquad \mathbf{v}_{a,b}(t) := \big\langle \mathbf{v}\big(a + t\,\tau^{a,b}\big), \tau^{a,b}\big\rangle \tag{5.59}$$

where $\mathbf{v} = \mathbf{i}^{-1}(\mathrm{id} - w)$, i.e. we shall operate with $\mathbf{v}_{a,b}$ instead of $\mathbf{w}_{a,b}$ as in Remark 5.36. It is clear that $\mathbf{v}_{a,b}$ is a maximal monotone map on $\mathbb{R}$. From the two point condition (5.53) written originally for $(u, \mathbf{v})$ we infer that

$$\frac{1}{2}\,|u_{a,b}(t_2) - u_{a,b}(t_1)|^2 \le \langle t_2' - t_1', t_2 - t_1\rangle \qquad \begin{array}{l}\forall\, t_1, t_2 \in \mathbb{R},\\ \forall\, t_1' \in \mathbf{v}_{a,b}(t_1),\ \forall\, t_2' \in \mathbf{v}_{a,b}(t_2).\end{array}$$

After Remark 5.36 we deduce that $u_{a,b} \in W^{1,2}(\mathbb{R})$ and there exists $v_{a,b} \in BV_{\mathrm{loc}} \cap L^\infty_{\mathrm{loc}}(\mathbb{R}; \mathbb{R})$ such that $\mathbf{v}_{a,b}(t) = \{v_{a,b}(t)\}$ for a.e. t and moreover

$$\frac{1}{2}\big(u_{a,b}'\big)^2 \le v_{a,b}' \qquad \text{a.e. in } \mathbb{R}.$$

where $u_{a,b}'$ is the weak derivative while $\mathrm{D}v_{a,b} = v_{a,b}'\,\mathcal{L}^1 + \mathrm{D}_s v_{a,b}$. Note that $\mathrm{D}_s v_{a,b} \in \mathcal{M}_+(\mathbb{R})$, see (5.54); clearly $\mathrm{D}v_{a,b} \ge 0$ as well. Thanks to Theorem 2.42 we can redefine $v_{a,b}$ on a null set so that it is monotone in the classical sense. Then, at each point $t \in \mathbb{R}$ the left and right sided limits $v_{a,b}(t^-)$ and $v_{a,b}(t^+)$ are well defined. Moreover it is straightforward to check that $\mathbf{v}_{a,b}(t) = [v_{a,b}(t^-), v_{a,b}(t^+)] \ni v_{a,b}(t)$ for any t.

The foregoing analysis of behaviour of $(u, w) \in \overline{\mathcal{X}}_1$ along straight lines makes it possible to characterize segments $[a, b]$ whose endpoints $x = a$, $y = b$ satisfy the two-point condition (5.50) as an equality. The result below will be of paramount importance in Chapter 6:

Proposition 5.37. *For functions $(u,w)\in\overline{\mathcal{X}}_1$ assume that there exists a pair of distinct points $(a,b)\in\overline{\Omega}\times\overline{\Omega}\setminus\Delta$ such that*

$$\frac{1}{2}\,|u(b)-u(a)|^2+\langle b'-a',b-a\rangle=|b-a|^2 \qquad \begin{array}{l}\forall\, a'\in\mathbf{w}(a),\\ \forall\, b'\in\mathbf{w}(b).\end{array} \tag{5.60}$$

where $\mathbf{w}=\mathbf{i}^{-1}(w)$. Then:

(i) u is affine on $[a,b]$;
(ii) $\big\langle\mathbf{w},\tau^{a,b}\big\rangle$ is univalued on $[a,b]$ (including the end-points) and it is affine there;
(iii) the two-point equality (5.60) holds true for any pair (x,y) satisfying $x,y\in[a,b]$.

Proof. With $\mathbf{v}=\mathrm{id}-\mathbf{w}$ we define $u_{a,b}$, $\mathbf{v}_{a,b}$ as in (5.59) and the associated univalued monotone function $v_{a,b}$. Since $u_{a,b}$ is an element of $W^{1,2}(\mathbb{R})$ it is absolutely continuous and therefore, by employing Schwarz inequality we obtain (we put $c:=|b-a|$):

$$\frac{1}{2}\big(u_{a,b}(c)-u_{a,b}(0)\big)^2=\frac{1}{2}\left(\int_{[0,c]}u'_{a,b}(t)\,dt\right)^2\le c\int_{[0,c]}\frac{1}{2}\big(u'_{a,b}(t)\big)^2\,dt\le c\int_{[0,c]}v'_{a,b}(t)\,dt$$
$$\le c\int_{[0,c]}v'_{a,b}(t)\,dt+c\int_{[0,c]}\mathrm{D}_s v_{a,b}=c\int_{[0,c]}\mathrm{D}v_{a,b}=c\,\big(v_{a,b}(c^+)-v_{a,b}(0^-)\big).$$

It is easy to check that the right hand side is equal to $|b-a|^2-\langle b'-a',b-a\rangle$ for some $a'\in\mathbf{w}(a)$, $b'\in\mathbf{w}(b)$ therefore from (5.60) we deduce that above we have equalities everywhere, which has several implications. First, since $\mathrm{D}_s v_{a,b}$ is a positive measure, there must hold $\mathrm{D}_s v_{a,b}=0$ on $[0,c]$ hence $\mathbf{v}_{a,b}$ is univalued and absolutely continuous therein. Next we observe that the Schwarz inequality becomes an equality therefore $u'_{a,b}$ is a.e. equal to a constant on $[0,c]$ furnishing the claim (i). From the chain of equalities we also infer that $\frac{1}{2}\big(u'_{a,b}\big)^2=v'_{a,b}$ a.e. on $[0,c]$ therefore $v_{a,b}$ is affine there as well, thus proving (ii). The statement (iii) is clear once the equality (5.60) is divided by $|b-a|^2$. □

5.4. Optimality conditions under extra regularity assumptions

In this section we are looking for a system of necessary and sufficient conditions for pairs (u,w) and (ϑ,ς) to be solutions of problems $(\overline{\mathscr{P}})$ and $(\mathscr{P}^*)$, respectively. From this moment on – in the remainder of the whole chapter – we will focus on the Canvas-like Material Design (CanMD) setting and we will confine ourselves to domains Ω that are star-shaped with respect to a ball – this way we have the characterization of the set $\overline{\mathcal{X}}_1$ stated in Theorem 5.32. We are thus looking at the pair of problems:

$$\max\left\{\int u\,df\ :\ (u,w)\in\overline{\mathcal{X}}_1\right\} \tag{$\overline{\mathscr{P}}$}$$

$$\min\left\{\int\varrho^0(\vartheta,\varsigma)\ :\ (\vartheta,\varsigma)\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d}_+),\ -\mathrm{div}\,\vartheta=f,\ \ \mathrm{Div}\,\varsigma=0\ \text{ in }\Omega\right\} \tag{$\mathscr{P}^*$}$$

where for $(q,\sigma)\in\mathbb{R}^d\times\mathcal{S}^{d\times d}_+$

$$\varrho^0(q,\sigma)=\mathrm{Tr}\,\sigma+\tfrac{1}{2}\langle\sigma^{-1}q,q\rangle,$$

cf. the formula (5.38) for ρ^0_+ in the CanMD setting.

Similarly as in Section 4.4 the optimality conditions shall employ the μ-tangential calculus summarized in Section 2.5. We will again exploit the fact that the pair $(\overline{\mathscr{P}})$, $(\mathscr{P}^*)$ is a particular case in the class of pairs $(\overline{\mathcal{P}})$, $(\mathcal{P}^*)$ studied in Chapter 4, which can be seen by taking $A(u,w) = \big(\nabla u, e(w)\big)$ and replacing ρ by ϱ. Sadly, the results from Section 4.4 cannot be repeated in full generality. While condition (B4) (formulated in the latter section) is fulfilled for the operator A, the condition (B5) is not, or more particularly: ϱ is not coercive. The effect, in a nutshell, is that there are pairs $(u,w) \in \overline{\mathcal{X}}_1$ which lie outside the domain $\mathrm{dom}(A_\mu)$.

Derivation of optimality conditions in the general case has turned out to be a very delicate issue and in this thesis we dodge resolving it. Instead we shall concentrate on a class of functions $(u,w) \in \overline{\mathcal{X}}_1$ that are Lipschitz continuous, which will guarantee that $(u,w) \in \mathrm{dom}(A_\mu)$. This simplified approach can be deemed reasonable since for convex domains Ω all the analytical and numerical solutions so far found (that will be presented in this chapter) enjoy this extra regularity property. We will return to discussing possible strategies for handling the general case in the open problem part, cf. Section 5.7.2.

5.4.1. The system of optimality conditions

As announced above our analysis will be limited to functions (u,w) that are Lipschitz continuous. Under such regularity assumption checking whether (u,w) is an element of the set $\overline{\mathcal{X}}_1$ simplifies, since in condition (iii) in Theorem 5.32 we have $\varepsilon_s(w) = 0$ while $e(w)$ can be a.e. computed in the classical sense owing to Rademacher theorem. We thus infer that:

Proposition 5.38. *A pair of functions $(u,w) \in \mathrm{Lip}_0(\Omega;\mathbb{R}^{d+1})$ is an element of the convex set $\overline{\mathcal{X}}_1$ if and only if*

$$\frac{1}{2}\nabla u \otimes \nabla u + e(w) \preceq \mathrm{I} \qquad \text{a.e. in } \Omega \tag{5.61}$$

where ∇u and $e(w) = \frac{1}{2}\big(\nabla w + (\nabla w)^\top\big)$ are indented in the classical sense.

In what follows we employ the μ-tangential calculus from Section 2.5 in the framework below:

$$\begin{cases} U = \mathbb{R}^{d+1}, \quad \mathscr{U} = \Omega, \quad \mathcal{U} = \mathcal{D}(\Omega;\mathbb{R}^{d+1}), \\ |||\cdot||| = \|\cdot\|_{\infty,\Omega} \quad \big(q = \infty\big), \\ V = \mathbb{R}^d \times \mathcal{S}^{d\times d}, \quad A = A(u,w) = \big(\nabla u, e(w)\big). \end{cases} \tag{5.62}$$

Remark 5.39. Note that the norm $|||\cdot|||$ chosen above for the μ-tangential analysis is different than the one employed for the relaxation of $(\mathscr{P})$, cf. (5.40). This is precisely why Section 4.4 cannot be exploited directly and we are obliged to make independent arguments as below.

Since in the setting (5.62) functions u and w are decoupled it is straightforward that for any $\mu \in \mathcal{M}^b_+(\Omega)$ we have

$$\mathrm{dom}(A_\mu) = \mathrm{dom}(\nabla_\mu) \times \mathrm{dom}(e_\mu), \qquad M_\mu = T_\mu \times M^e_\mu, \qquad A_\mu(u,w) = \big(\nabla_\mu u, e_\mu(w)\big)$$

for any $u \in \operatorname{dom}(\nabla_\mu) \subset C_0(\Omega)$ and $w \in \operatorname{dom}(e_\mu) \subset C_0(\Omega;\mathbb{R}^d)$, see Sections 2.5.2 and 2.5.3. Next, for $\mu \in \mathcal{M}^b_+(\Omega)$ assume $(q,\sigma) \in L^1_\mu(\Omega;\mathbb{R}^d \times \mathcal{S}^{d\times d})$ such that: $q(x) \in T_\mu(x)$, $\sigma(x) \in M^e_\mu(x)$ for μ-a.e. x and $-\operatorname{div}(q\mu) = f$ and $-\operatorname{Div}(\sigma\mu) = 0$. The two equations can be written as $A^*(q\mu,\sigma\mu) = \mathscr{F}$ for $\mathscr{F}(u,w) = \int u\,df$, therefore from Theorem 2.57 we get that for any $(u,w) \in \operatorname{dom}(A_\mu)$ there holds the integration by parts formula:

$$\int_\Omega \Big(\langle \nabla_\mu u, q\rangle + \langle e_\mu(w), \sigma\rangle\Big) d\mu = \int_\Omega u\,df. \tag{5.63}$$

Based on Proposition 2.65 we now show that any Lipschitz continuous pair (u,w) that vanishes on $\partial\Omega$ belongs in fact to the domain of operator A_μ:

$$\operatorname{Lip}_0(\Omega;\mathbb{R}^{d+1}) \subset \big(\operatorname{dom}(\nabla_\mu)\big)^{d+1} \subset \operatorname{dom}(\nabla_\mu) \times \operatorname{dom}(e_\mu) = \operatorname{dom}(A_\mu),$$

where to write down the second inclusion we additionally acknowledged Proposition 2.66.

Before stating the optimality conditions we require a technical result that brings the properties of the set $\overline{\mathcal{X}}_1 \cap \operatorname{Lip}_0(\Omega;\mathbb{R}^{d+1})$ to the μ-tangential setting (note that the proof of the lemma includes an independent proof that this set is contained in $\operatorname{dom}(A_\mu)$):

Lemma 5.40. *For $\mu \in \mathcal{M}^b_+(\Omega)$ let us take any functions $(q,\sigma) \in L^1_\mu(\overline{\Omega};\mathbb{R}^d \times \mathcal{S}^{d\times d}_+)$ such that $q \in T_\mu$ and $\sigma \in M^e_\mu$ μ-a.e. Then, for any $(u,w) \in \overline{\mathcal{X}}_1 \cap \operatorname{Lip}_0(\Omega;\mathbb{R}^{d+1})$*

$$\langle \nabla_\mu u, q\rangle + \langle e_\mu(w), \sigma\rangle \le \varrho^0(q,\sigma) = \operatorname{Tr}\sigma + \tfrac{1}{2}\langle \sigma^{-1}q, q\rangle \qquad \mu\text{-a.e. in } \Omega. \tag{5.64}$$

while the equality holds if and only if

$$\Big\langle \frac{1}{2}\,\nabla_\mu u \otimes \nabla_\mu u + e_\mu(w), \sigma \Big\rangle = \operatorname{Tr}\sigma \qquad \text{and} \qquad q = \sigma\,\nabla_\mu u \qquad \mu\text{-a.e. in } \Omega. \tag{5.65}$$

Proof. For fixed $(u,w) \in \operatorname{Lip}_0(\Omega;\mathbb{R}^{d+1})$ we must point to a sequence $(u_n,w_n) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1})$ such that:

(1) $(u_n,w_n) \to (u,w)$ uniformly in Ω,
(2) $\sup_n \|\big(\nabla u_n, e(w_n)\big)\|_{\Omega,\infty} < \infty$,
(3) $\varrho\big(\nabla u_n, e(w_n)\big) \le 1$ in Ω for $n \in \mathbb{N}$.

With such a sequence at our disposal for every Borel set $B \subset \Omega$ and for every n by exploiting the property (3) we obtain

$$\int_B \Big(\langle \nabla u_n, q\rangle + \langle e(w_n), \sigma\rangle\Big) d\mu \le \int_B \varrho^0(q,\sigma)\,d\mu.$$

From the property (2) we can see that, up to extracting a subsequence, $\big(\nabla u_n, e(w_n)\big) \overset{*}{\rightharpoonup} (\theta,\epsilon)$ in $L^\infty_\mu(\Omega;\mathbb{R}^d \times \mathcal{S}^{d\times d})$. Since $q \in T_\mu$, $\sigma \in M^e_\mu$ and $(u,w) \in \operatorname{dom}(A_\mu)$ there must hold $\langle \theta, q\rangle + \langle \epsilon, \sigma\rangle = \langle P_\mu\theta, q\rangle + \big\langle \Pi^e_\mu\,\epsilon, \sigma\big\rangle = \langle \nabla_\mu u, q\rangle + \langle e_\mu(w), \sigma\rangle$ μ-a.e., see Definition 2.51. As a result, by passing with n to infinity in the inequality above we infer that

$$\int_B \Big(\langle \nabla_\mu u, q\rangle + \langle e_\mu(w), \sigma\rangle\Big) d\mu \le \int_B \varrho^0(q,\sigma)\,d\mu$$

thus (5.64) follows by localizing owing to arbitrariness of the Borel set $B \subset \Omega$.

Therefore inequality (5.64) will be proved if we are able to find a sequence (u_n,w_n) satisfying (1), (2), (3). By definition of ϱ the condition (3) is equivalent to $(u_n,w_n) \in \mathcal{X}_1$. Since $(u,w) \in \overline{\mathcal{X}}_1$ the sequence may put forth by repeating word for word the construction

used in the proof of implication (iii) $\Rightarrow$ (i) in Theorem 5.32. Indeed, by standard properties of mollifying the extra regularity $(u,w)\in \mathrm{Lip}_0(\Omega;\mathbb{R}^{d+1})$ guarantees the uniform convergence of (u_n,w_n) and that $\mathrm{lip}(u_n)\le \mathrm{lip}(u)$, $\mathrm{lip}(w_n)\le \mathrm{lip}(w)$. Inequality (5.64) is now established.

We turn to examine the conditions under which (5.64) is an equality. First we observe that the inequality (5.64) may be μ-a.e. in Ω rewritten as

$$\left\langle \tfrac{1}{2}\nabla_\mu u\otimes\nabla_\mu u+e_\mu(w),\sigma\right\rangle+\langle\nabla_\mu u,q\rangle\le \mathrm{Tr}\,\sigma+j_2(\sigma,\nabla_\mu u)+j_2^*(\sigma,q) \tag{5.66}$$

where we recall that $j_2(\sigma,\nabla_\mu u)=\frac{1}{2}\langle\sigma\,\nabla_\mu u,\nabla_\mu u\rangle$ and $j_2^*(\sigma,q)=\frac{1}{2}\langle\sigma^{-1}q,q\rangle$. We know that (5.66) holds for any pair of (q,σ) satisfying the prerequisite of the lemma. In particular we can test with $(q,\sigma)=(\sigma\nabla_\mu u,\sigma)$ where $\sigma\in L^1_\mu(\overline{\Omega};\mathcal{S}^{d\times d}_+)$ with $\sigma\in M^e_\mu$ μ-a.e. Indeed, by Proposition 2.66 we have $\mathrm{Im}(\sigma)\subset T_\mu$. For such pair the inequality (5.66) reduces to $\left\langle \frac{1}{2}\nabla_\mu u\otimes\nabla_\mu u+e_\mu(w),\sigma\right\rangle\le \mathrm{Tr}\,\sigma$, while clearly $\langle\nabla_\mu u,q\rangle\le j_2(\sigma,\nabla_\mu u)+j_2^*(\sigma,q)$ and the latter inequality becomes equality if and only if $q=\sigma\,\nabla_\mu u$. All in all we deduce that (5.66) is an equality if and only if conditions (5.65) hold true. □

The optimality conditions for the pair of problems $(\overline{\mathscr{P}})$, $(\mathscr{P}^*)$ confined to the case of Lipschitz continuous functions (u,w) may readily be formulated:

Theorem 5.41. *Let us take a quintuple (u,w,μ,σ,q) such that: $(u,w)\in \mathrm{Lip}_0(\Omega;\mathbb{R}^{d+1})$, $\mu\in\mathcal{M}_+(\overline{\Omega})$, $\sigma\in L^1_\mu(\overline{\Omega};\mathcal{S}^{d\times d}_+)$, $q\in L^1_\mu(\overline{\Omega};\mathbb{R}^d)$.*

Then, the pairs (u,w) and $(\vartheta,\varsigma)=(q\mu,\sigma\mu)$ solve problems $(\overline{\mathscr{P}})$ and $(\mathscr{P}^*)$, respectively, if and only if the following optimality conditions are met:

$$\begin{cases} (i) & -\mathrm{div}(q\mu)=f,\quad \mathrm{Div}(\sigma\mu)=0 \quad \text{in } \Omega,\\ (ii) & \frac{1}{2}\nabla u\otimes\nabla u+e(w)\preceq \mathrm{I}\quad \text{a.e. in } \Omega,\\ (iii) & \mu(\partial\Omega)=0,\\ (iv) & \left\langle \frac{1}{2}\nabla_\mu u\otimes\nabla_\mu u+e_\mu(w),\sigma\right\rangle=\mathrm{Tr}\,\sigma \quad \mu\text{-a.e.},\\ (v) & q=\sigma\,\nabla_\mu u \quad \mu\text{-a.e.} \end{cases} \tag{5.67}$$

Proof. With the inequality (5.64) at our disposal we may repeat the arguments from the proof of Theorem 4.20 with the operator $A=A(u,w)=\big(\nabla u,e(w)\big)$, the functional $\mathscr{F}=\mathscr{F}(u,w)=\int u\,df$ and the gauge function ϱ in the place of ρ (see also Remark 4.21). The optimality condition from Theorem 4.20 then would read: (i') $A^*(q\mu,\sigma\mu)=(f,0)$; (ii') $(u,w)\in\overline{\mathcal{X}}_1$; (iii) $\mu(\partial\Omega)=0$; (iv') $\langle(A_\mu(u,w),(q,\sigma)\rangle=\varrho^0(q,\sigma)$ μ-a.e. As discussed above Lemma 5.40 condition (i') translates directly to (i). Thanks to Proposition 5.38 conditions (ii) and (ii') are equivalent as well. Equivalence between (iv') and the pair of conditions (iv), (v) is the precise statement of the "moreover part" in Lemma 5.40. □

Once we obtain a solution (u,w,μ,σ,q) of the system (5.67) (with the additional normalization $\mathrm{Tr}\,\sigma=1$ μ-a.e.) we may utilize Corollary 5.19 to find a solution λ to (OEM) problem where we can exploit the characterization (5.39) of the set $\mathcal{K}_{1,+}\big(\sigma(x)\big)=\mathcal{K}_1\big(\sigma(x)\big)$. Instead we may follow Section 4.4 and adapt the optimality conditions (5.67) towards

the (OEM) problem by involving the Hooke tensor field $\mathscr{H}$, see Theorem 4.22 and its Corollary 4.24. This passage requires invoking the notion of μ-tangential energies:

$$j_{+,\mu}(x,H,\xi)=\inf_{\zeta\in N^e_\mu(x)} j_+(H,\xi+\zeta),\qquad \rho_{+,\mu}(\xi)=\inf_{\zeta\in N^e_\mu(x)}\rho_+(\xi+\zeta), \tag{5.68}$$

where $N^e_\mu(x)=\big(M^e_\mu(x)\big)^\perp$, see Section 4.4 for details.

Proposition 5.42. *For any $(u,w)\in\overline{\mathcal{X}}_1\cap \mathrm{Lip}_0(\Omega;\mathbb{R}^{d+1})$ we have*

$$\rho_{+,\mu}\Big(x\,,\,\tfrac12\nabla_\mu u\otimes\nabla_\mu u+e_\mu(w)\Big)\le 1\qquad \mu\text{-a.e.} \tag{5.69}$$

Proof. From the proof of Lemma 5.40 we learned that for μ-a.e. x we have $\big\langle\big(\frac12\nabla_\mu u\otimes\nabla_\mu u+e_\mu(w)\big)(x),\sigma\big\rangle\le\operatorname{Tr}\sigma$ for every $\sigma\in\mathcal{S}^{d\times d}_+\cap M^e_\mu(x)$. Since $\rho^0_{+,\mu}(x,\tilde\sigma)=\operatorname{Tr}\tilde\sigma$ when $\tilde\sigma\in\mathcal{S}^{d\times d}_+\cap M^e_\mu(x)$ and $\rho^0_{+,\mu}(x,\tilde\sigma)=\infty$ otherwise, we actually infer that $\big\langle\big(\frac12\nabla_\mu u\otimes\nabla_\mu u+e_\mu(w)\big)(x),\tilde\sigma\big\rangle\le\rho^0_{+,\mu}(x,\tilde\sigma)$ for any $\tilde\sigma\in\mathcal{S}^{d\times d}$, which proves the assertion by definition of polarity. □

Corollary 5.43. *Let us take a 6-tuple $(u,w,\mu,\sigma,q,\mathscr{H})$ such that: $(u,w)\in\mathrm{Lip}_0(\Omega;\mathbb{R}^{d+1})$, $\mu\in\mathcal{M}_+(\overline\Omega)$, $\sigma\in L^1_\mu(\overline\Omega;\mathcal{S}^{d\times d}_+)$, $q\in L^1_\mu(\overline\Omega;\mathbb{R}^d)$, $\mathscr{H}\in L^\infty_\mu(\overline\Omega;\mathcal{K}_{fib})$ and $\operatorname{Tr}\mathscr{H}\le 1$ μ-a.e. Then the following conditions:*

(1) $\lambda=\mathscr{H}\mu$ solves the (OEM) *problem in the CanMD setting,*
(2) (u,w) is a relaxed solution of the maximization problem (5.9)*,*
(3) (q,σ) solves the minimization problem (5.13) *for decomposition $\lambda=\mathscr{H}\mu$*

hold true if and only if the 6-tuple solves the system below:

$$\begin{cases}
(i) & -\mathrm{div}(q\mu)=f,\quad \mathrm{Div}(\sigma\mu)=0 \quad \text{in } \Omega,\\
(ii) & \frac12\nabla u\otimes\nabla u+e(w)\preceq\big(\frac{\mathcal{Z}}{2C_0}\big)^{\frac23}\mathrm{I}\quad \text{a.e. in } \Omega,\\
(iii) & \mu(\partial\Omega)=0,\\
(iv) & \sigma\in\partial j_{+,\mu}\big(x\,,\,\mathscr{H}\,,\,\frac12\nabla_\mu u\otimes\nabla_\mu u+e_\mu(w)\big)\quad \mu\text{-a.e.},\\
(v) & q=\sigma\,\nabla_\mu u\quad \mu\text{-a.e.}\\
(vi) & \operatorname{Tr}\sigma=\big(\frac{\mathcal{Z}}{2C_0}\big)^{\frac23}\quad \mu\text{-a.e.}
\end{cases} \tag{5.70}$$

Proof. First let us assume that $\frac{\mathcal{Z}}{2C_0}=1$. Then, considering Theorem 5.17 and Proposition 5.28 (note that $p=2$ here) the set of conditions $(1),(2),(3)$ may be rewritten as: (1') (u,w) solves $(\overline{\mathscr{P}})$, (2') $(q\mu,\sigma\mu)$ solves $(\mathscr{P}^*)$, (3') $\mathscr{H}(x)\in\underline{\mathcal{K}}_1\big(\sigma(x)\big)$ for μ-a.e. x, (4') $\operatorname{Tr}\sigma=1$ μ-a.e. Then, owing to Theorem 5.41, (1'), (2') are equivalent to (i)-(v) in (5.67). Since $(u,w)\in\overline{\mathcal{X}}_1$ thanks to Proposition 5.42 the triple: $\xi=\big(\frac12\nabla u\otimes\nabla u+e(w)\big)(x)$, $\sigma(x)$, $\breve{H}=\mathscr{H}(x)$ satisfies the prerequisite of Corollary 4.16 for μ-a.e. x (note that $\rho^0_+(\sigma)=\operatorname{Tr}\sigma=1$). Readily the pair of conditions (iv) in 5.67, (3') can be rewritten as (iv) in (5.70). To sum up we have showed that $(1),(2),(3)$ are equivalent to (5.70) if $\frac{\mathcal{Z}}{2C_0}=1$. The claim in the general case follows by revisiting Theorem 5.17 and Proposition 5.28 and by performing simple rescaling. □

Remark 5.44. We observe that the system of conditions above incorporates the μ-tangential variant of the PDE system governing the statics of elastic membrane – compare conditions (i), (iv), (v) in (5.70) with (5.14).

5.4.2. Analytical solutions in the radial-symmetric case

We shall now exploit the optimality conditions derived above in order to give analytical solutions to the pair of problems $(\overline{\mathscr{P}})$, $(\mathscr{P})$ in the radial symmetric case. For $R > 0$ we consider a circular domain $\Omega = \{x \in \mathbb{R}^2 : |x| < R\}$ and any radial symmetric load f (note that f can be a signed measure). The load is then determined by the bounded repartition function

$$\mathcal{F}(r) \;:=\; \int_{\{|x|\le r\}} f(dx), \qquad r \in [0,R]. \tag{5.71}$$

Then, working in polar coordinates (r,θ) and the corresponding local basis $\mathrm{e}_r, \mathrm{e}_\theta$, we will show that there always exists a solution (u,w,μ,σ,q) to the system (5.67) in the form (we put $r = |x|$):

$$u(x) = \tilde{u}(r), \quad w(x) = w_r(r)\,\mathrm{e}_r, \quad \mu = \mathcal{L}^2 \llcorner \Omega, \quad \sigma = \sigma_{rr}(r)\,\mathrm{e}_r \otimes \mathrm{e}_r, \quad q = q_r(r)\,\mathrm{e}_r \tag{5.72}$$

for suitable $\tilde{u}, w_r \in \mathrm{Lip}\big((0,R)\big)$ and q_r, σ_{rr} in $L^1_{\mathrm{loc}}\big((0,R)\big)$ where $\sigma_{rr} \ge 0$. We can use the condition (v) in (5.67) to deduce rightaway that $q_r(r) = \sigma_{rr}(r)\,u'(r)$. Thus the quintuple of the form (5.72) solves the system of optimality conditions if and only if:

$$\begin{cases} (0) & \tilde{u}(R) = 0, \quad w_r(0) = w_r(R) = 0, \\ (i') & -\mathrm{div}\big(\sigma_{rr}\,\tilde{u}'\,\mathrm{e}_r\big) = f, \quad \mathrm{Div}\big(\sigma_{rr}\,\mathrm{e}_r \otimes \mathrm{e}_r\big) = 0 \quad \text{in } \Omega, \\ (ii') & \frac{1}{2}\,|\tilde{u}'|^2 + w_r' \le 1, \quad \frac{w_r}{r} \le 1 \qquad \text{a.e. in } (0,R), \\ (iv') & \frac{1}{2}\,|\tilde{u}'|^2 + w_r' = 1 \qquad \text{a.e. in } \{\sigma_{rr} > 0\}, \end{cases} \tag{5.73}$$

where in (0) we use (i) and express $(u,w) = (0,0)$ on $\partial\Omega$ and moreover that $w = w_r(|x|)\,\mathrm{e}_r$ is continuous at $x = 0$. The conditions (ii'), (iv') are equivalent to (ii), (iv) since

$$\frac{1}{2}\nabla u \otimes \nabla u + e(w) = \Big(\frac{1}{2}\,|\tilde{u}'|^2 + w_r'\Big)\mathrm{e}_r \otimes \mathrm{e}_r + \frac{w_r}{r}\,\mathrm{e}_\theta \otimes \mathrm{e}_\theta, \qquad \mathrm{Tr}\,\sigma = \sigma_{rr}(r).$$

It turns out that (5.73) admits a unique solution. Indeed, solving the equations in (i') leads to

$$\sigma_{rr}(r) = \frac{D}{r}, \qquad \tilde{u}'(r) = -\frac{\mathcal{F}(r)}{2\pi D} \tag{5.74}$$

for a suitable positive constant D. This can be checked by noticing that σ_{rr} needs to satisfy $\sigma_{rr}' + \sigma_{rr}/r = 0$ in $(0,R)$ and that conversely the distributional divergence of $\frac{1}{r}\,\mathrm{e}_r \otimes \mathrm{e}_r$ vanishes on the whole disk Ω (including the origin). On the other hand, the expression for $u'(r)$ is deduced by integrating the first equation of (i') on the subset $\{|x| < r\}$ and by using Green's formula and (5.71). Accordingly, in order to match with (0) and with the equality constraint in (iv'), we deduce that:

$$\tilde{u}(r) = \frac{1}{2\pi D}\int_r^R \mathcal{F}(t)\,dt, \qquad w_r(r) = r - \frac{1}{8\pi^2 D^2}\int_0^r \big(\mathcal{F}(t)\big)^2\,dt \tag{5.75}$$

where the constant D is determined by the condition $w_r(R) = 0$, i.e.

$$D = \frac{1}{2\pi}\left(\frac{1}{2R}\int_0^R \big(\mathcal{F}(t)\big)^2\, dt\right)^{\frac{1}{2}}. \tag{5.76}$$

Note that $\tilde{u}$ and w_r are both Lipschitz (since $|\mathcal{F}(t)| \le \int |f|$) and that the first inequality in (ii') is an equality while the second one (i.e. $w \le r$) follows directly from (5.75). Therefore all the required conditions in (5.73) are satisfied.

We will now specify the obtained solutions for two radial-symmetric loads f:

Example 5.45. For the disk domain $\Omega = \big\{x \in \mathbb{R}^2 \,:\, |x| < R\big\}$ we consider a uniformly distributed load $f = p\,\mathcal{L}^2 \llcorner \Omega$ that acts upwards on the membrane (p is a positive constant in pascals). After computing $\mathcal{F} = \mathcal{F}(r)$ we obtain from (5.74), (5.75), (5.76) that

$$\tilde{u}(r) = \frac{\sqrt{10}}{3}\frac{R^3 - r^3}{R^2}, \quad w_r(r) = r - \frac{r^5}{R^4}, \quad q_r(r) = -\frac{p\,r}{2}, \quad \sigma_{rr}(r) = \frac{pR^2}{2\sqrt{10}}\,\frac{1}{r} \tag{5.77}$$

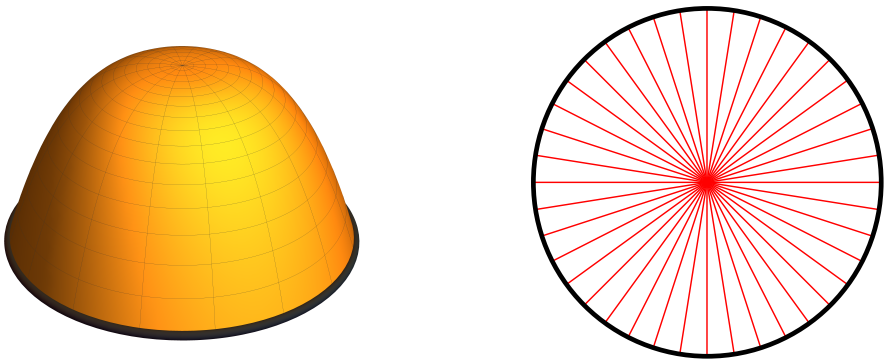
where in addition we acknowledged that $q_r = \sigma_{rr}\,\tilde{u}'$. The graph of function $u = u(x) = \tilde{u}(|x|)$ is showed in Fig. 5.2(a). One can show that the pre-stress field $\varsigma = \sigma_{rr}\mathrm{e}_r \otimes \mathrm{e}_r\,\mathcal{L}^2 \llcorner \Omega$ decomposes to infinite family of strings coinciding with diameters of the disk Ω, more precisely:

$$\langle \epsilon, \varsigma \rangle_{(C,\mathcal{M})} = \frac{pR^2}{\sqrt{10}} \int_{(0,\pi)} \left(\int \langle \epsilon, \varsigma^{x(\theta),y(\theta)} \rangle \right) d\theta$$

where $x(\theta) := (R\cos\theta, R\sin\theta)$ and $y(\theta) := -x(\theta)$. It now becomes more clear that $\operatorname{Div}\varsigma = 0$ in $\mathcal{D}'(\Omega;\mathbb{R}^2)$. Fig. 5.2(b) reflects the fibrous character of the field ς. The value function $\mathcal{Z}$ may be equivalently computed as $\max \overline{\mathscr{P}}$ or $\min \mathscr{P}^*$:

$$\mathcal{Z} = \int_\Omega u\, df = \int_\Omega \varrho^0(q,\sigma)\, d\mathcal{L}^2 = \frac{2\pi}{\sqrt{10}}\, pR^3.$$

△



(a) (b)

Figure 5.2: Solution to a radial-symmetric optimal membrane problem for a uniformly distributed load $f = p\,\mathcal{L}^2 \llcorner \Omega$: (a) the graph of function u solving $(\overline{\mathscr{P}})$; (b) the pre-stress field ς solving $(\mathscr{P}^*)$ that decomposes to strings going from boundary to boundary.

As was already mentioned the system (5.73) has a unique solution, but it is not necessarily the case for the more general system (5.67). Indeed, we passed to system (5.73) by making a restrictive assumption that σ is of the rank-one form $\sigma = \sigma_{rr}(r)\,\mathrm{e}_r \otimes \mathrm{e}_r$, while the additional term $\sigma_{\theta\theta}(r)\,\mathrm{e}_\theta \otimes \mathrm{e}_\theta$ may also occur in general. That being said, in Example 5.45 one can show that there must hold $\sigma_{\theta\theta} = 0$ after all, since $\langle \mathcal{A}(u,w), \mathrm{e}_\theta \otimes \mathrm{e}_\theta \rangle = \frac{w_r}{r} = 1 - (r/R)^4 < 1$. Below we give another radial-symmetric example where apart from ς of the form visible in Fig. 5.2(a) we have infinitely many other possibilities.

Example 5.46. For Ω being again a disk of radius R that is centred at the origin we consider a load $f = p\,\mathcal{L}^2 \llcorner \left\{R > |x| > \frac{R}{2}\right\}$, i.e. the load is uniformly distributed in the annulus of the inner radius $\frac{R}{2}$. From (5.74), (5.75), (5.76) we obtain that the quadruple

$$\tilde{u}(r) = \begin{cases} 2\sqrt{\frac{10}{57}}\,R & \text{if } r \le \frac{R}{2}, \\ \sqrt{\frac{10}{57}}\,\frac{1}{R^2}\left(R^3 + 3R^2 r - 4\,r^3\right) & \text{if } r \ge \frac{R}{2}, \end{cases}$$

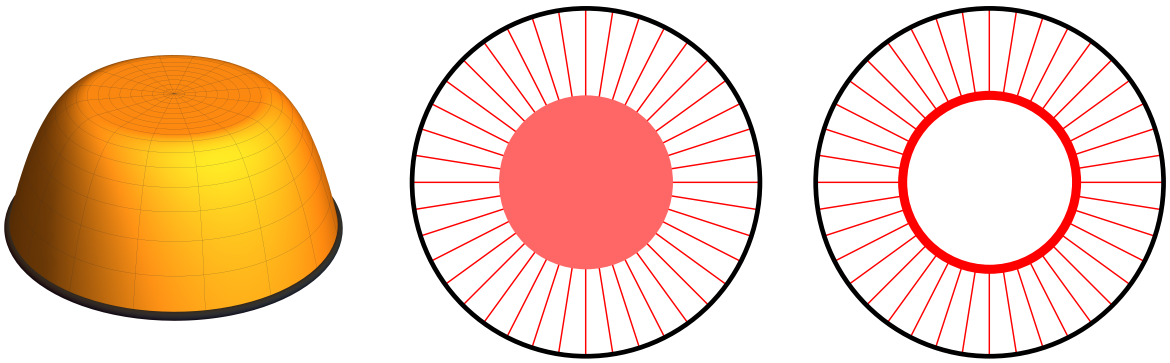
$$w_r(r) = \begin{cases} r & \text{if } r \le \frac{R}{2}, \\ \frac{4}{19R^4}\left(R^5 + R^4 r + 10R^2 r^3 - 12\,r^5\right) & \text{if } r \ge \frac{R}{2}, \end{cases} \tag{5.78}$$

$$q_r(r) = \begin{cases} 0 & \text{if } r \le \frac{R}{2}, \\ p\left(\frac{R^2}{8\,r} - \frac{r}{2}\right) & \text{if } r \ge \frac{R}{2}, \end{cases} \qquad \sigma_{rr}(r) = \sqrt{\frac{19}{30}}\,\frac{pR^2}{8}\,\frac{1}{r}$$

solves the system (5.73) and the value function reads

$$\mathcal{Z} = \int_\Omega \varrho^0(\vartheta, \varsigma) = \int_\Omega u\,f = \frac{\pi}{2}\sqrt{\frac{19}{30}}\,pR^3.$$

Fig. 5.3 presents the graph of the function $u = u(x) = \tilde{u}(|x|)$, while the pre-stress field $\varsigma = \sigma_{rr}\,\mathrm{e}_r \otimes \mathrm{e}_r\,\mathcal{L}^2 \llcorner \Omega$ is again represented by infinitely many fibres lying along the diameters, cf. Fig. 5.2(b).



(a) (b) (c)

Figure 5.3: Solution to a radial-symmetric optimal membrane problem for the load f distributed in the annulus $\left\{R > |x| > \frac{R}{2}\right\}$: (a) the graph of function u solving $(\overline{\mathscr{P}})$; (b) alternative optimal pre-stress field ς_2 with a hydrostatic part in the inner disk; (c) another optimal pre-stress field ς_2 with a 1D ring in tension.

We will now point to other solutions ς of the problem $(\mathscr{P}^*)$. We observe that function u admits a plateau over the smaller disk $\{|x| \leq \frac{R}{2}\}$, while $w = w_r \mathrm{e}_r$ is equal to identity map therein. Consequently, $\nabla u = 0$ in this disk (note that functions u, w are smooth) while $\frac{1}{2}\nabla u \otimes \nabla u + e(w) = e(w) = \mathrm{I}$, thus to solve the more general system (5.67) of optimality conditions, $\varsigma \llcorner \{|x| \leq \frac{R}{2}\}$ may be modified while minding the in-plane equilibrium $\mathrm{Div}\,\varsigma = 0$ only. Indeed, conditions (iv), (v) in (5.67) are automatically met. Accordingly, we propose two examples of alternative solutions ς_1, ς_2 to problem $(\mathscr{P}^*)$:

$$\varsigma_1 = \sqrt{\frac{19}{30}}\frac{pR^2}{8}\left(\frac{1}{r}\,\mathrm{e}_r \otimes \mathrm{e}_r\,\mathcal{L}^2 \llcorner \{|x| > \tfrac{R}{2}\} + \frac{2}{R}\,\mathrm{I}\,\mathcal{L}^2 \llcorner \{|x| < \tfrac{R}{2}\}\right)$$

and

$$\varsigma_2 = \sqrt{\frac{19}{30}}\frac{pR^2}{8}\left(\frac{1}{r}\,\mathrm{e}_r \otimes \mathrm{e}_r\,\mathcal{L}^2 \llcorner \{|x| > \tfrac{R}{2}\} + \mathrm{e}_\theta \otimes \mathrm{e}_\theta\,\mathcal{H}^1 \llcorner \{|x| = \tfrac{R}{2}\}\right).$$

The pre-stress field $\varsigma_1 = \sigma_1\,\mathcal{L}^2 \llcorner \Omega$ is showed in Fig. 5.3(b). We can see that in the inner disk $\{|x| < \frac{R}{2}\}$ the function σ_1 is of rank two and hydrostatic. The equilibrium condition $\mathrm{Div}\,\varsigma_1 = 0$ can be easily verified by using Green's formula independently on the inner disk and the annulus. With $q_1 = q_r\,\mathrm{e}_r$ it is straightforward to check that the quintuple $(u, w, \mathcal{L}^2 \llcorner \Omega, \sigma_1, q_1)$ solves the system (5.67).

The field ς_2 vanishes in the open inner disk $\{|x| < \frac{R}{2}\}$ whilst the equilibrium is assured by a 1D ring in tension, see Fig. 5.3(c). To show that actually $\mathrm{Div}\,\varsigma_2 = 0$ in Ω one can use Green's formula for the annulus, while by definition of distributional divergence we obtain $-\mathrm{Div}\big(\mathrm{e}_\theta \otimes \mathrm{e}_\theta\,\mathcal{H}^1 \llcorner \{|x| = r_0\}\big) = \frac{1}{r_0}\,\mathrm{e}_r\,\mathcal{H}^1 \llcorner \{|x| = r_0\}$ in $\mathcal{D}'(\Omega;\mathbb{R}^2)$ for any $r_0 \in (0, R)$. In order to validate optimality conditions (5.67) we can propose a decomposition $\varsigma_2 = \sigma_2\,\mu_2$ where $\mu_2 = \mu_{2,a} + \mu_{2,s}$ with $\mu_{2,a} = \mathcal{L}^2 \llcorner \{R > |x| > \frac{R}{2}\}$ and $\mu_{2,s} = \mathcal{H}^1 \llcorner \{|x| = \frac{R}{2}\}$. Then, by virtue of Propositions 2.59 and 2.67 the μ_2-tangential spaces (with respect to operator ∇ and e respectively) read

$$T_{\mu_2} = \begin{cases} \mathbb{R}^2 & \mu_{2,a}\text{-a.e.,} \\ \mathrm{span}(\mathrm{e}_\theta) & \mu_{2,s}\text{-a.e.,} \end{cases} \qquad M^e_{\mu_2} = \begin{cases} \mathcal{S}^{2\times 2} & \mu_{2,a}\text{-a.e.,} \\ \mathrm{span}(\mathrm{e}_\theta \otimes \mathrm{e}_\theta) & \mu_{2,s}\text{-a.e.} \end{cases}$$

thus furnishing

$$\big(\nabla_{\mu_2} u, e_{\mu_2}(w)\big) = \begin{cases} \big(\tilde{u}'(r)\,\mathrm{e}_r,\ w_r'(r)\,\mathrm{e}_r \otimes \mathrm{e}_r + \frac{w_r(r)}{r}\,\mathrm{e}_\theta \otimes \mathrm{e}_\theta\big) & \mu_{2,a}\text{-a.e.,} \\ \big(0,\ \frac{w_r(R/2)}{R/2}\,\mathrm{e}_\theta \otimes \mathrm{e}_\theta\big) & \mu_{2,s}\text{-a.e.} \end{cases}$$

By acknowledging that

$$\sigma_2 = \begin{cases} \sqrt{\frac{19}{30}}\,\frac{pR^2}{8}\,\frac{1}{r}\,\mathrm{e}_r \otimes \mathrm{e}_r & \mu_{2,a}\text{-a.e.,} \\ \sqrt{\frac{19}{30}}\,\frac{pR^2}{8}\,\mathrm{e}_\theta \otimes \mathrm{e}_\theta & \mu_{2,s}\text{-a.e.} \end{cases} \qquad q_2 = \sigma_2 \nabla_{\mu_2} u = \begin{cases} p\left(\frac{R^2}{8r} - \frac{r}{2}\right)\mathrm{e}_r & \mu_{2,a}\text{-a.e.,} \\ 0 & \mu_{2,s}\text{-a.e.} \end{cases}$$

one can readily verify that optimality conditions (5.67) are met for the quintuple $(u, w, \mu_2, \sigma_2, q_2)$. Let us stress that in this example the use of μ-tangential calculus was purely illustrative since ∇u and $e(w)$ could be used in (5.67) by smoothness of (u, w). This would not be the case for the knife load $f = p\,\mathcal{H}^1 \llcorner \{|x| = \frac{R}{2}\}$ where (up to multiplicative constants) ς_1 and ς_2 would still be optimal and, although optimal u and w would be

non-differentiable on $\{|x| = \frac{R}{2}\}$, the μ-tangential computations as above would do the job. $\triangle$

5.5. Optimal membranes that are systems of strings. The numerical method by ground structure approach

In Section 4.7 we have recalled an alternative perspective on the Michell problem (MP) when domain Ω is convex – a formulation that originally falls within the family of Free Material Design problems and thus may be written as a pair $(\mathcal{P}), (\mathcal{P}^*)$. From Theorem 4.55 we have found that one can concentrate on the problem (4.64) where a truss (a system of straight bars) is being looked for. A truss is represented through a signed pairing measure $\mathscr{S} \in \mathcal{M}(\overline{\Omega} \times \overline{\Omega}; \mathbb{R})$. For most loading and boundary conditions scenarios, however, the truss problem (4.64) does not admit solutions. Nonetheless any minimizing sequence of trusses $\mathscr{S}_n$ furnishes a minimizing sequence $\varsigma_{\mathscr{S}_n}$ (see the integration formula (4.58)) for the (MP) problem. Furthermore, restricting the support of $\mathscr{S}$ to $\mathrm{X}^h \times \mathrm{X}^h$ (X^h is a finite subset of $\overline{\Omega}$) paves the way towards the famous numerical scheme by the so-called *ground structure approach*, cf. Section 4.7.3 for details.

For Michell problem the underlying optimal truss design was to be expected from the mechanical point of view – after all Michell structures were originally invented as limit points of minimizing sequences of trusses. This cannot be told about the membrane problem (OEM) or (OPM). On the other hand, mathematically speaking, the truss problem (4.64) can be recast directly from Michell problem by an alternative duality scheme that employs the two-point constraint that characterizes the set of admissible displacement functions $\overline{\mathcal{U}}_1$, see Theorem 4.55. At this point we know that in the CanMD setting of the (OEM) problem we have a similar characterization: the pair of continuous functions $(u, w) \in C_0(\Omega; \mathbb{R}^{d+1})$ is an element of the set $\overline{\mathcal{X}}_1$ if and only if

$$\frac{1}{2}\,|u(y) - u(x)|^2 + \langle w(y) - w(x), y - x\rangle \leq |y - x|^2 \qquad \forall\,(x, y) \in \overline{\Omega} \times \overline{\Omega}, \tag{5.79}$$

which is by virtue of Theorem 5.32 combined with Remark 5.34.

Under assumption of convexity of Ω in this section we use the two-point condition (5.79) to establish a competing duality scheme between the problem $(\mathscr{P})$ and a new problem $(\widetilde{\mathscr{P}}^*)$ where we look for an optimal *system of strings*. We distinguish string systems from planar trusses in the classical sense: both structures are composed of straight segments, yet trusses act within a plane in which they lie while string systems work out of the plane. Unfortunately, similarly as in the case of the optimal truss problem, solution to $(\widetilde{\mathscr{P}}^*)$ will fail to exist in general. Notwithstanding this it seems that there is a wide class of data f for which $(\widetilde{\mathscr{P}}^*)$ admits a solution, which will be demonstrated through numerical simulations, see also the conjecture in Section 5.7.1 below. Finally, we will develop the numerical method for the optimal membrane problem via ground structure approach – we will look for an optimal string system among those spanned by a finite set of nodes X^h.

5.5.1. The alternative duality scheme using the two-point condition

Throughout the whole Section 5.5 we will assume that

the domain Ω is convex.

Let us depart by introducing a closed (unbounded) convex subset of $\mathbb{R}^2$:

$$\mathrm{K} := \left\{ (t_1, t_2) \in \mathbb{R}^2 \,:\, \frac{1}{2}\,(t_1)^2 + t_2 \,\leq 1 \right\} \tag{5.80}$$

which is directly related with the two-point condition (5.42). To show this we define two linear continuous operators $H : C_0(\Omega;\mathbb{R}) \to C(\overline{\Omega}\times\overline{\Omega};\mathbb{R})$ and $\mathcal{B} : C_0(\Omega;\mathbb{R}^d) \to C(\overline{\Omega}\times\overline{\Omega};\mathbb{R})$ such that for distinct pairs $(x,y) \in \overline{\Omega}\times\overline{\Omega}\setminus\Delta$ we have (recall that $\Delta = \big\{(x,x) : x \in \mathbb{R}^d\big\}$ is the diagonal)

$$(Hu)(x,y) := u(y) - u(x), \qquad (\mathcal{B}w)(x,y) := \left\langle w(y) - w(x), \frac{y-x}{|y-x|} \right\rangle \tag{5.81}$$

while, by convention, $(Hu)(x,y) = 0$ and $(\mathcal{B}u)(x,y) = 0$ whenever $x = y$. It is typical to call H and $\mathcal{B}$ *geometric operators*. Then, for any continuous pair $(u,w) \in C_0(\Omega;\mathbb{R}^{d+1})$ there holds an equivalence

$$(u,w) \in \overline{\mathcal{X}}_1 \qquad \Leftrightarrow \qquad \big(Hu, \mathcal{B}w\big)(x,y) \in |y-x|\,\mathrm{K} \qquad \forall\,(x,y) \in \overline{\Omega}\times\overline{\Omega}. \tag{5.82}$$

Indeed, for $x \neq y$ to the set $|y-x|\,\mathrm{K}$ belong those pairs (t_1,t_2) which satisfy $\frac{1}{2}\left(\frac{t_1}{|y-x|}\right)^2 + \frac{t_2}{|y-x|} \leq 1$ therefore by putting $t_1 = (Hu)(x,y)$, $t_2 = (\mathcal{B}w)(x,y)$ we recover the two-point condition (5.79).

The support function of the set K is closely related to the one of C:

Lemma 5.47. *The support function of* K *is a one-homogeneous convex lower semi-continuous non-negative function given by*

$$\chi^*_{\mathrm{K}}(s_1,s_2) = \begin{cases} \big(1 + \frac{1}{2}\,a^2\big)\,s_2 & \text{if } \ s_2 \geq 0 \ \text{ and } \ s_1 = a\,s_2, \\ +\infty & \text{if } \ s_2 < 0 \ \text{ or if } \ s_2 = 0,\ s_1 \neq 0. \end{cases} \tag{5.83}$$

It satisfies the relation

$$\chi^*_{\mathrm{K}}(s_1,s_2) = \chi^*_{\mathrm{C}}(s_1\,\tau,\ s_2\,\tau\otimes\tau) = \varrho^0(s_1\,\tau,\ s_2\,\tau\otimes\tau) \qquad \forall\,\tau \in S^{d-1}. \tag{5.84}$$

where C *is defined in* (5.21).

Proof. The formula (5.83) follows directly from (5.25) once we have shown (5.84). For every $(\theta,\xi) \in \mathrm{C}$ and any $\tau \in S^{d-1}$ we have $\frac{1}{2}\big(\langle\theta,\tau\rangle\big)^2 + \langle\xi,\tau\otimes\tau\rangle = \big\langle \frac{1}{2}\,\theta\otimes\theta + \xi, \tau\otimes\tau\big\rangle \leq 1$, thus $(t_1,t_2) = \big(\langle\theta,\tau\rangle, \langle\xi,\tau\otimes\tau\rangle\big)$ belongs to K and therefore

$$\chi^*_{\mathrm{K}}(s_1,s_2) \ \geq\ \sup\Big\{ s_1\,\langle\theta,\tau\rangle + s_2\,\langle\xi,\tau\otimes\tau\rangle \ :\ (\theta,\xi) \in \mathrm{C} \Big\} = \chi^*_{\mathrm{C}}(s_1\,\tau,\ s_2\,\tau\otimes\tau).$$

To obtain the opposite inequality, we associate to any element $(t_1,t_2) \in \mathrm{K}$ the pair $(\theta,\xi) \in \mathrm{C}$ defined by $\theta = t_1\,\tau$ and $\xi = t_2\,\tau\otimes\tau$. We are led to

$$\chi^*_{\mathrm{C}}(s_1\,\tau, s_2\,\tau\otimes\tau) \ \geq\ \sup\Big\{ t_1\,s_1 + t_2\,s_2 : (t_1,t_2) \in \mathrm{K} \Big\} = \chi^*_{\mathrm{K}}(s_1,s_2),$$

which establishes the first equality in (5.84), while the second one is due to equality $\chi^*_{\mathrm{C}} = \varrho^0$, see (5.23). □

Similarly as in Sections 4.7 and 4.8 for distinct pair of points $(x,y)\in\overline{\Omega}\times\overline{\Omega}\setminus\Delta$ we introduce 1D measures $(\vartheta^{x,y},\varsigma^{x,y})\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d}_+)$:

$$\vartheta^{x,y}:=\tau^{x,y}\,\mathcal{H}^1\llcorner[x,y],\qquad \varsigma^{x,y}:=\tau^{x,y}\otimes\tau^{x,y}\,\mathcal{H}^1\llcorner[x,y],\qquad \tau^{x,y}:=\frac{y-x}{|y-x|},$$

while for $x=y$ we set $\vartheta^{x,y}=0$, $\varsigma^{x,y}=0$ and $\tau^{x,y}=0$ by convention. The measure $\varsigma^{x,y}$ represents a string that connects points x and y and that is subject to a unit tensile pre-stressing force. On the other hand $\vartheta^{x,y}$ reflects a string carrying a unit transverse force. Ultimately a *string system* will be identified by two pairing measures:

$$\pi\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R})\quad\text{and}\quad\mathscr{S}\in\mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$$

where π and $\mathscr{S}$ describe distributions of transverse and, respectively, pre-stressing forces on the family of strings connecting all pairs of points in $\overline{\Omega}$. The transverse force field $\vartheta_\pi\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$ and the pre-stress field $\varsigma_{\mathscr{S}}\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)$ related to this string system may be constructed by integrating measures $\vartheta^{x,y}$ and $\varsigma^{x,y}$ respectively, i.e.

$$\langle\phi,\vartheta_\pi\rangle_{(C,\mathcal{M})}=\int_{\overline{\Omega}\times\overline{\Omega}}\left(\int\langle\phi,\vartheta^{x,y}\rangle\right)\pi(dxdy)\qquad\forall\,\phi\in C(\overline{\Omega};\mathbb{R}^d),\tag{5.85}$$

$$\langle\psi,\varsigma_{\mathscr{S}}\rangle_{(C,\mathcal{M})}=\int_{\overline{\Omega}\times\overline{\Omega}}\left(\int\langle\psi,\varsigma^{x,y}\rangle\right)\mathscr{S}(dxdy)\qquad\forall\,\psi\in C(\overline{\Omega};\mathcal{S}^{d\times d}).\tag{5.86}$$

Let us emphasise that the fact that $\operatorname{spt}(\vartheta_\pi,\varsigma_{\mathscr{S}})$ is contained in $\overline{\Omega}$ is due to the assumed convexity of domain Ω.

In the spirit of Section 4.7 the differential equilibrium equations $-\operatorname{div}\vartheta_\pi=f$, $-\operatorname{Div}\varsigma_{\mathscr{S}}=0$ in Ω may be equivalently written as equilibrium equations of joints that tie the string system. The linear continuous operators $H,\mathcal{B}$ admit their adjoints: $H^*:\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R})\to\mathcal{M}^b(\Omega;\mathbb{R})$ and $\mathcal{B}^*:\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R})\to\mathcal{M}^b(\Omega;\mathbb{R}^d)$. Then we have the following equivalences:

$$-\operatorname{div}\vartheta_\pi=f\quad\text{in }\mathcal{D}'(\Omega)\qquad\Leftrightarrow\qquad H^*\pi=f\quad\text{in }\mathcal{M}^b(\Omega),\tag{5.87}$$

$$-\operatorname{Div}\varsigma_{\mathscr{S}}=0\quad\text{in }\mathcal{D}'(\Omega;\mathbb{R}^d)\qquad\Leftrightarrow\qquad\mathcal{B}^*\mathscr{S}=0\quad\text{in }\mathcal{M}^b(\Omega;\mathbb{R}^d).\tag{5.88}$$

The equivalence (5.88) is the statement of Proposition (4.54) for $F=0$ and $\Upsilon_0=\partial\Omega$. The proof of (5.87) is similar and may be obtained through a simple modification of the proof of Proposition 4.62. The right hand sides of (5.87) and (5.88) express the out-of-plane and, respectively, in-plane equilibrium of points in Ω being potential joints of the string system $(\pi,\mathscr{S})$. For convenience we introduce the *set of statically admissible string systems*:

$$\Xi(\Omega,f):=\Big\{(\pi,\mathscr{S})\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)\,:\,H^*\pi=f,\ \mathcal{B}^*\mathscr{S}=0,\ \mathscr{S}\geq0\Big\}.$$

Directly from definition of the adjoint operator we obtain a characterization of the set $\Xi(\Omega,f)$: for a pair of measures $(\pi,\mathscr{S})\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$ we have an equivalence

$$(\pi,\mathscr{S})\in\Xi(\Omega,f)\quad\Leftrightarrow\quad\begin{cases}(i)\ \int_{\overline{\Omega}\times\overline{\Omega}}\big(\varphi(y)-\varphi(x)\big)\,\pi(dxdy)=\int_{\overline{\Omega}}\varphi\,df & \forall\,\varphi\in C_0(\Omega)\\ (ii)\ \int_{\overline{\Omega}\times\overline{\Omega}}\langle\phi(y)-\phi(x),\tau^{x,y}\rangle\,\mathscr{S}(dxdy)=0 & \forall\,\phi\in C_0(\Omega;\mathbb{R}^d)\\ (iii)\ \mathscr{S}\geq0\end{cases}\tag{5.89}$$

Following the Goffman-Serrin construction (see Section 2.2), we may associate to any $(\pi, \mathscr{S}) \in \mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$ the non-negative Borel measure $\chi^*_{\mathrm{K}}(\pi,\mathscr{S})$ ranging in $[0,+\infty]$. The functional $\mathcal{J} : \mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2) \to \overline{\mathbb{R}}$ defined as below

$$\mathcal{J}(\pi,\mathscr{S}) := \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\ \chi^*_{\mathrm{K}}\big(\pi,\mathscr{S}\big)(dxdy) \tag{5.90}$$

will play a role of the total cost (or total energy) of a string system $(\pi,\mathscr{S})$.

Lemma 5.48. *The functional $\mathcal{J}$ enjoys the following properties:*

(i) $\mathcal{J}$ is convex, weakly- lower semi-continuous and 1-homogeneous on $\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$ and it is the support function of the non-empty closed convex set*

$$\mathscr{K} := \Big\{(v_1,v_2)\in C(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)\ :\ \big(v_1(x,y),v_2(x,y)\big)\in |y-x|\,\mathrm{K}\quad \forall\,(x,y)\in\overline{\Omega}\times\overline{\Omega}\Big\}.$$

(ii) In order that $\mathcal{J}(\pi,\mathscr{S})<+\infty$, it is necessary that $\mathscr{S}\geq 0$ and $\pi \ll \mathscr{S}$ on $\overline{\Omega}\times\overline{\Omega}\setminus\Delta$. In this case, it holds that

$$\mathcal{J}(\pi,\mathscr{S}) = \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\left(1+\frac{1}{2}\left(\frac{d\pi}{d\mathscr{S}}\right)^2\right)\mathscr{S}(dxdy). \tag{5.91}$$

(iii) For every $(\pi,\mathscr{S})\in\mathcal{M}(\overline{\Omega};\mathbb{R}^2)$ we have the inequality

$$\int_{\overline{\Omega}} \varrho^0(\vartheta_\pi,\varsigma_{\mathscr{S}})\ \leq\ \mathcal{J}(\pi,\mathscr{S}).$$

Proof. Let us define a multifunction $\Gamma:\overline{\Omega}\to 2^{\mathbb{R}^2}$: for any $(x,y)\in\overline{\Omega}\times\overline{\Omega}$ we put $\Gamma(x,y) = |y-x|\,\mathrm{K}$. It is easy to check that Γ satisfies the prerequisites of Theorem 2.23 for $X=\overline{\Omega}\times\overline{\Omega}$, $V=\mathbb{R}^2$ and $\mathscr{C}=\mathscr{K}$, in particular Γ is lower semi-continuous, including the diagonal Δ where $\Gamma=\{0\}$. It is straightforward to check that $\chi^*_{\Gamma(x,y)}(s_1,s_2) = |y-x|\,\chi^*_{\mathrm{K}}(s_1,s_2)$. By equality (2.12) in Theorem 2.23 we find that $\mathcal{J}$ is precisely the support function of $\mathscr{K}$ from which follows the weak-* lower semi-continuity. The statement (i) is proved.

We now turn to show (ii). For given $(\pi,\mathscr{S})$ let us assume that $\mathcal{J}(\pi,\mathscr{S})$ is finite. Then, by a standard use of Fubini's theorem

$$\mathcal{J}(\pi,\mathscr{S}) = \int_0^\infty \chi^*_{\mathrm{K}}(\pi,\mathscr{S})(B_t)\,dt$$

where $B_t=\{|x-y|>t\}$. Further, if $\pi=\alpha\,\mathscr{S}+\pi_s$ with $\pi_s\perp\mathscr{S}$ and $\alpha\in L^1_{\mathscr{S}}(\overline{\Omega}\times\overline{\Omega})$ is the Lebesgue-Nikodym decomposition of π with respect to $\mathscr{S}$, we have $\chi^*_{\mathrm{K}}(\pi,\mathscr{S}) = \chi^*_{\mathrm{K}}(\alpha\,\mathscr{S},\mathscr{S}) + \chi^*_{\mathrm{K}}(\pi_s,0)$, hence by Lemma 5.47 for every $t>0$:

$$\chi^*_{\mathrm{K}}(\pi,\mathscr{S})(B_t) = \begin{cases} \int(1+\frac{\alpha^2}{2})\,d\mathscr{S} & \text{if } \ \mathscr{S}\llcorner B_t\geq 0 \ \text{ and } \ |\pi_s|(B_t)=0,\\ +\infty & \text{otherwise.}\end{cases}$$

Since $\cup_{t>0}B_t = \overline{\Omega}\times\overline{\Omega}\setminus\Delta$, by integrating with respect to t between 0 and $+\infty$, we recover the necessary conditions required for the finiteness of $\mathcal{J}(\pi,\mathscr{S})$. Under these conditions the formula (5.91) holds true by Goffman-Serrin convention and again by Lemma 5.47.

Let us now prove (iii). It is not restrictive to assume that $\mathcal{J}(\pi,\mathscr{S}) < +\infty$ so that $\mathscr{S} \geq 0$ and $\pi = \alpha\mathscr{S}$ on $\overline{\Omega}\times\overline{\Omega}\setminus\Delta$ for a suitable Borel function α. Then, for any continuous selection (ϕ,ψ) of the convex set C (see (5.21)), we have

$$\begin{aligned}
\langle \vartheta_\pi, \phi\rangle + \langle \varsigma_{\mathscr{S}}, \psi\rangle &= \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} \Big(\alpha(x,y)\langle \phi, \vartheta^{x,y}\rangle + \langle \psi, \varsigma^{x,y}\rangle\Big)\, \mathscr{S}(dxdy)\\
&= \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} \left(\int_{[x,y]} \Big(\alpha(x,y)\,\langle \phi, \tau^{x,y}\rangle + \langle \psi, \tau^{x,y}\otimes\tau^{x,y}\rangle\Big) d\mathcal{H}^1\right) \mathscr{S}(dxdy)\\
&\leq \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} \left(\int_{[x,y]} \chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big)\, d\mathcal{H}^1\right) \mathscr{S}(dxdy)\\
&= \int_{\overline{\Omega}\times\overline{\Omega}\setminus\Delta} |x-y|\ \chi^*_{\mathrm{K}}(\pi,\mathscr{S})(dxdy) \leq \mathcal{J}(\pi,\mathscr{S})
\end{aligned}$$

where, to pass from second to the third line, we used the fact that $\big(\phi(z),\psi(z)\big)$ belongs to the convex set C for all $z\in[x,y]$ so that, in virtue of (5.84), we have

$$\langle \phi(z), \alpha(x,y)\,\tau^{x,y}\rangle + \langle \psi(z), \tau^{x,y}\otimes\tau^{x,y}\rangle \leq \varrho^0\big(\alpha(x,y)\tau^{x,y}, \tau^{x,y}\otimes\tau^{x,y}\big) = \chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big).$$

We conclude with the desired inequality by taking the supremum of the left hand side above with respect to all pairs (ϕ,ψ) being continuous selections of the set C, see (5.28). □

We may readily formulate the problem of designing an optimal string system:

$$\inf\Big\{\mathcal{J}(\pi,\mathscr{S})\ :\ (\pi,\mathscr{S})\in\Xi(\Omega,f)\Big\} \tag{$\widetilde{\mathscr{P}}^*$}$$

Remark 5.49. The equilibrium conditions (i), (ii) in (5.89) *a priori* allows signed measures $(\pi,\mathscr{S})$. However, in order to ensure that $\varsigma_{\mathscr{S}}$ is positive definite one has to additionally impose condition (iii) on $\mathscr{S}$. In fact the infimum in $(\widetilde{\mathscr{P}}^*)$ is unchanged if we restrict ourselves to measures π of the kind $\pi=\alpha\mathscr{S}$ with $\alpha\geq 0$ or, as will appear more natural in the context of Section 5.6, with $\alpha\leq 0$. Indeed, if $(\alpha\,\mathscr{S},\mathscr{S})$ is an element of the admissible set $\Xi(\Omega,f)$, then so is $(|\alpha|\,\mathscr{S}_1,\mathscr{S}_1)$ where $\mathscr{S}_1$ is defined by $\langle\varphi,\mathscr{S}_1\rangle_{(C,\mathcal{M})} := \int_{\alpha\geq 0}\varphi(x,y)\,d\mathscr{S} + \int_{\alpha<0}\varphi(y,x)\,d\mathscr{S}$ for every $\varphi\in C(\overline{\Omega}\times\overline{\Omega})$. Clearly this new admissible pair shares the same energy i.e. $\mathcal{J}(|\alpha|\,\mathscr{S}_1,\mathscr{S}_1) = \mathcal{J}(\alpha\,\mathscr{S},\mathscr{S})$. This will also be the case for the admissible pair $(-|\alpha|\,\mathscr{S}_2,\mathscr{S}_2)$ where $\mathscr{S}_2$ is the image of $\mathscr{S}_1$ under the map $(x,y)\mapsto(y,x)$.

The main result of this subsection views the problem $(\widetilde{\mathscr{P}}^*)$ as a dual of $(\overline{\mathscr{P}})$ (with zero duality gap) that is alternative to $(\mathscr{P}^*)$. Its proof will require several lemmas given below.

Theorem 5.50. *The following chain of zero-gap equalities holds true:*

$$\mathcal{Z} = \max\overline{\mathscr{P}} = \min\mathscr{P}^* = \inf\widetilde{\mathscr{P}}^*.$$

Moreover any minimizing sequence $(\pi_n,\mathscr{S}_n)$ for the problem $(\widetilde{\mathscr{P}}^*)$ furnishes a minimizing sequence $(\vartheta_{\pi_n},\varsigma_{\mathscr{S}_n})$ for the problem $(\mathscr{P}^*)$. In particular, if there exists a solution $(\pi,\mathscr{S})$ of $(\widetilde{\mathscr{P}}^*)$ then $(\vartheta_\pi,\varsigma_{\mathscr{S}})$ is a solution of $(\mathscr{P}^*)$.

Remark 5.51. In contrast to problem $(\mathscr{P}^*)$ a solution to problem $(\widetilde{\mathscr{P}^*})$ is not guaranteed to exist in general. This failure of existence is due to the fact that minimizing sequence $(\pi_n, \mathscr{S}_n) = (\alpha_n \mathscr{S}_n, \mathscr{S}_n)$, despite the natural estimate $\sup_n \mathcal{J}(\pi_n, \mathscr{S}_n) = \sup_n \int (1+\frac{\alpha_n^2}{2})\,|y-x|\,\mathscr{S}_n(dxdy) < \infty$, can exhibit large mass concentrations of $\mathscr{S}_n$ near the diagonal Δ. The same kind of difficulty arises in the mathematical approach towards Michell's truss problem developed in [42]. In Remark 5.59 we shall show that for data in Example 5.45 an optimal string system cannot be constructed, although Fig. 5.2(b) displaying ς may suggest the contrary.

Remark 5.52. Through the whole Section 5.5 we keep the assumption on convexity of the domain Ω. One of the obvious reasons for this is the construction of measures $(\vartheta_\pi, \varsigma_\mathscr{S})$ from the string system $(\pi, \mathscr{S}) \in \Xi(\Omega, f)$, see (5.85), (5.86). Clearly spt $(\vartheta_\pi, \varsigma_\mathscr{S})$ is not contained in $\overline{\Omega}$ once spt $(\vartheta_\pi, \varsigma_\mathscr{S})$ contains pairs of points (x, y) such that $[x, y] \not\subseteq \overline{\Omega}$. Then again, such a pair $(\pi, \mathscr{S})$ will never be optimal for $(\widetilde{\mathscr{P}^*})$ owing to the fact that whole $\partial\Omega$ is a Dirichlet zone. To understand why this is so, let us consider a simple pair $(\pi, \mathscr{S}) = (\alpha\,\delta_{(x,y)}, \delta_{(x,y)})$ for $\alpha \in \mathbb{R}$ such that $[x, y] \cap \overline{\Omega} = [x, x_0] \cup [y_0, y]$ where $x_0, y_0 \in \partial\Omega$ and $|y_0 - x_0| > 0$. We put $\widetilde{\mathscr{S}} = \delta_{(x,x_0)} + \delta_{(y_0,y)}$, $\widetilde{\pi} = \alpha\,\widetilde{\mathscr{S}}$ and then clearly

$$\mathcal{J}(\widetilde{\pi}, \widetilde{\mathscr{S}}) = |x_0 - x|\,\big(1 + \tfrac{\alpha^2}{2}\big) + |y - y_0|\,\big(1 + \tfrac{\alpha^2}{2}\big) < |y - x|\,\big(1 + \tfrac{\alpha^2}{2}\big) = \mathcal{J}(\pi, \mathscr{S}).$$

Meanwhile $H\widetilde{\pi} = H\pi$ and $\mathcal{B}\widetilde{\mathscr{S}} = \mathcal{B}\mathscr{S}$ owing to the fact that $x_0, y_0 \in \partial\Omega$. We therefore have a consistent way of rebuilding $(\pi, \mathscr{S}) \in \Xi(\Omega, f)$ so that the induced measures $(\vartheta_\pi, \varsigma_\mathscr{S})$ are in fact supported in $\overline{\Omega}$. We also note that the two-point characterization of the set $\mathcal{X}_1$ in Lemma 5.22 (or its non-smooth version $\overline{\mathcal{X}}_1$ in Theorem 5.32) does not necessitate convexity of Ω. Ultimately one should keep in mind that most of this section could be adapted to non-convex design domains. Such generalization, however, would greatly complicate the formalism used in this section thus we shall stick to the convexity assumption after all. A numerical simulation for a non-convex Ω will be given in Example 5.75.

The alternative duality scheme, that from problem $(\mathscr{P})$ yields the dual $(\widetilde{\mathscr{P}^*})$, will lead through the two-point characterization of the set $\overline{\mathcal{X}}_1$ (more precisely its smooth version $\mathcal{X}_1$) according to (5.82). This time around the duality argument will be more involved as Theorem 2.12 cannot be used: for $\Lambda(u, w) := \big(Hu, \mathcal{B}w\big)$ the functional $\Psi : C(\overline{\Omega} \times \overline{\Omega}) \to \overline{\mathbb{R}}$, being the indicator function of the set of continuous selection of the multifunction $(x, y) \mapsto |y - x|\,\mathrm{K}$, is not continuous at any $\Lambda(u_0, w_0)$. The issue concerns the neighbourhood of the diagonal Δ where $|y - x|\,\mathrm{K}$ degenerates to a set that does not contain a neighbourhood of the origin in $\mathbb{R}^2$.

Instead of relying on Theorem 2.12 we shall perform the duality argument "more manually", that is by considering the following perturbation of problem $(\overline{\mathscr{P}})$: for G being any closed subset of $\overline{\Omega} \times \overline{\Omega}$ we define $h_G : C(\overline{\Omega} \times \overline{\Omega}; \mathbb{R}^2) \to \overline{\mathbb{R}}$ to be the functional given by

$$h_G(p, q) := \inf_{u,w} \left\{ -\int u\,df \;\middle|\; \begin{array}{l} (u, w) \in C_0(\Omega; \mathbb{R}^{d+1}), \\ \big(Hu + p,\, \mathcal{B}w + q\big)(x, y) \in |y - x|\,\mathrm{K} \quad \forall\,(x, y) \in G \end{array} \right\}. \tag{5.92}$$

To prove Theorem 5.50 we will only require the perturbation function for $G = \overline{\Omega} \times \overline{\Omega}$, for which we denote $h := h_{\overline{\Omega}\times\overline{\Omega}}$. The duality analysis for other sets G will prove useful in later sections. By equivalence (5.82) we can see that the set of admissible pairs (u, w) above contains $\mathcal{X}_1$ and is contained in $\overline{\mathcal{X}}_1$ (we have $\mathcal{D}(\Omega;\mathbb{R}^{d+1}) \subset C_0(\Omega;\mathbb{R}^{d+1}) \subset C_0(\Omega) \times \mathcal{W}_\Omega$) and as a result

$$h(0,0) = -\max \overline{\mathscr{P}} = -\mathcal{Z}. \tag{5.93}$$

On the path towards proving Theorem 5.50 we need additional technical results:

Lemma 5.53. *Let there be given two sequences: $(p_n, q_n) \in C(\overline{\Omega} \times \overline{\Omega};\mathbb{R}^2)$ and $(u_n, w_n) \in C_0(\Omega;\mathbb{R}^{d+1})$ such that $(p_n, q_n) \to (0,0)$ uniformly while for every n we have*

$$\frac{1}{2}\big|(Hu_n + p_n)(x,y)\big|^2 + |y-x|\,(\mathcal{B}w_n + q_n)(x,y) \le |y-x|^2 \qquad \forall (x,y) \in \overline{\Omega} \times \overline{\Omega}. \tag{5.94}$$

Then there exist a subsequence u_{n_k} and a pair $(u, w) \in \overline{\mathcal{X}}_1$ such that $u_{n_k} \to u$ uniformly on $\overline{\Omega}$.

Proof. It is not restrictive to assume that $|p_n(x,y)| \le 1$ and $|q_n(x,y)| \le 1$ for any n and $(x,y) \in \overline{\Omega} \times \overline{\Omega}$. Recalling definition (5.81) we see that

$$\langle w_n(y) - w_n(x), y - x\rangle \le |x-y|^2 - q_n(x,y)|y-x| \le |y-x|(|y-x|+1).$$

From this we deduce that w_n is uniformly bounded by slightly modifying the proof of claim (i) in Lemma 5.24. Indeed, by choosing for any $x \in \Omega$, any vector $\tau \in S^{d-1}$ and a real t such that $|t| \le \operatorname{diam}(\Omega)$ and $y = x + t\tau \in \partial\Omega$, we obtain that $-\langle w(x), \tau\rangle \le t + 1 \le D + 1$ where $D = \operatorname{diam}(\Omega)$. As τ is arbitrary, we arrive at the uniform bound $\|w_n\|_{\Omega,\infty} \le D + 1$ yielding in particular that $\mathcal{B}w_n + q_n \ge -2(D+1) - 1$. Going back to the inequality (5.94), we obtain the following estimate on u_n:

$$|u_n(y) - u_n(x) + p_n(x,y)| \le C\sqrt{|y-x|} \qquad \forall (x,y) \in \overline{\Omega} \times \overline{\Omega}, \tag{5.95}$$

where we can take $C = \big(6D+6\big)^{1/2}$. Since p_n tends to zero uniformly in $\overline{\Omega}$ we infer equi-continuity of $\{u_n\}$ and, since $u_n \subset C_0(\Omega)$, in virtue of Arzelà-Ascoli theorem, we may extract a subsequence u_{n_k} that converges uniformly to a function $u \in C_0(\Omega)$.

Next, for each k we define a univalued multifunction $\mathbf{w}_k : \overline{\Omega} \to 2^{\mathbb{R}^d}$ by putting $\mathbf{w}_k(x) = \{w_{n_k}(x)\}$. Then each graph $\mathcal{G}_{\mathbf{w}_k}$ is a closed subset of the compact set $\overline{\Omega}\times\overline{B}(0, D+1)$ where the closed ball is a subset of $\mathbb{R}^d$. According to Proposition 2.29, up to extracting a subsequence (without further relabelling), $\mathcal{G}_{\mathbf{w}_k}$ converge in the sense of Kuratowski to a closed set $\mathcal{G}$. Let $\tilde{\mathbf{w}} : \overline{\Omega} \to 2^{\mathbb{R}^d}$ be the multifunction whose graph is $\mathcal{G}$; note that $\mathbf{w}_k \to \tilde{\mathbf{w}}$ in the sense of graph, see (2.16). Then $\mathbf{w}(x) \ne \varnothing$ for every $x \in \overline{\Omega}$. Indeed, since $w_{n_k}(x) \in \overline{B}(0, D+1)$ the sequence $\big(x, w_{n_k}(x)\big)$ admits a cluster point (x, x') and therefore $x' \in \tilde{\mathbf{w}}(x)$ by the first property in (2.16). Since $\mathcal{G}$ is closed and $\tilde{\mathbf{w}}$ has values in the subsets of a compact set $\overline{B}(0, D+1)$, owing to Theorem 2.31 the multifunction $\tilde{\mathbf{w}}$ is upper semi-continuous. We extend $\tilde{\mathbf{w}}$ (we do not relabel) to a multifunction on whole $\mathbb{R}^d$ by defining $\tilde{\mathbf{w}}(x) = \{0\}$ for $x \in \mathbb{R}^d \setminus \overline{\Omega}$. The upper semi-continuity is preserved as $0 \in \tilde{\mathbf{w}}$ on $\partial\Omega$ due to $w_{n_k} \in C_0(\Omega;\mathbb{R}^d)$. To sum up: $\mathbf{w}_k = \{w_{n_k}\}$ (also extended by $\{0\}$ to $\mathbb{R}^d$) converge in the sense of graph to $\tilde{\mathbf{w}} : \mathbb{R}^d \to 2^{\mathbb{R}^d}$ being an upper semi-continuous multifunction of full domain.

Next, for an arbitrary pair $(x,y)\in\overline{\Omega}\times\overline{\Omega}$ we choose any $x'\in\tilde{\mathbf{w}}(x)$ and $y'\in\tilde{\mathbf{w}}(y)$. Then, owing to the second property of graph convergence in (2.16), there exists a sequence $(x_k,y_k)\in\overline{\Omega}\times\overline{\Omega}$ such that $(x_k,y_k)\to(x,y)$ and $\big(w_{n_k}(x_{n_k}),w_{n_k}(y_{n_k})\big)\to(x',y')$. Directly from (5.94) we have

$$\tfrac{1}{2}\big(u_{n_k}(y_k)-u_{n_k}(x_k)+p_{n_k}(x_k,y_k)\big)^2+\langle w_{n_k}(y_k)-w_{n_k}(x_k),y_k-x_k\rangle \\ +q_{n_k}(x_k,y_k)\,|y_k-x_k|\le|y_k-x_k|^2.$$

Since u_n, p_n and q_n converge uniformly on $\overline{\Omega}$ we find that in the limit

$$\tfrac{1}{2}\,|u(y)-u(x)|^2+\langle y'-x',y-x\rangle\le|y-x|^2\qquad\forall(x,y)\in\overline{\Omega}\times\overline{\Omega},\quad\forall\,\hat{x}\in\tilde{\mathbf{w}}(x),\ \hat{y}\in\tilde{\mathbf{w}}(y) \tag{5.96}$$

rendering $\tilde{\mathbf{v}}:=\mathrm{id}-\tilde{\mathbf{w}}$ an upper semi-continuous monotone multifunction of full domain. In case $\tilde{\mathbf{v}}$ is not maximal we define $\mathbf{v}(x)=\mathrm{co}\big(\tilde{\mathbf{v}}(x)\big)$ arriving at a maximal monotone map $\mathbf{v}\supset\tilde{\mathbf{v}}$ according to the criterion in Proposition 2.40 ($\mathbf{v}$ inherits the upper semi-continuity property due to the finite dimension of the domain). We infer that $\mathbf{w}=\mathrm{id}-\mathbf{v}$ is an element of $\mathbf{W}_\Omega$. It is easy to check that (5.96) holds true with $\tilde{\mathbf{w}}$ replaced by $\mathbf{w}$ and therefore, for $w=\mathbf{i}(\mathbf{w})$, $(u,w)\in\overline{\mathcal{X}}_1$ by Theorem 5.32, see the criterion (ii). The proof is complete. □

Lemma 5.54. *Let G be a compact subset of $\overline{\Omega}\times\overline{\Omega}$ and define the (possibly empty) convex set*

$$\Xi_G(\Omega,f):=\Big\{(\pi,\mathscr{S})\in\Xi(\Omega,f)\ :\ \mathrm{spt}\big((\pi,\mathscr{S})\big)\subset G\Big\}.$$

Then:

(i) The conjugate of h_G in the duality between $C(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$ and $\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$ is given by

$$h_G^*(\pi,\mathscr{S})=\begin{cases}\mathcal{J}(\pi,\mathscr{S}) & \text{if } (\pi,\mathscr{S})\in\Xi_G(\Omega,f),\\ +\infty & \text{otherwise.}\end{cases}$$

(ii) h_G is convex; it is finite at $(0,0)$ if and only if $\Xi_G(\Omega,f)\cap\mathrm{dom}(\mathcal{J})$ is non-empty. If this is the case and if moreover G does not intersect the diagonal Δ, then h_G is continuous at $(0,0)$ and there exists a minimizer for the problem

$$\inf\Big\{\mathcal{J}(\pi,\mathscr{S})\ :\ (\pi,\mathscr{S})\in\Xi_G(\Omega,f)\Big\}\qquad(\widetilde{\mathscr{P}}_G^*)$$

while the zero-gap equality holds: $\min\widetilde{\mathscr{P}}_G^*=-h_G(0,0)$.

(iii) The function $h:=h_{\overline{\Omega}\times\overline{\Omega}}$ is finite and lower semi-continuous at $(0,0)$.

Proof. Let us compute the conjugate of h_G at $(\pi,\mathscr{S})$. We have

$$h_G^*(\pi,\mathscr{S})=\sup_{p,q,u,w}\left\{\langle p,\pi\rangle+\langle q,\mathscr{S}\rangle+\langle u,f\rangle\ \middle|\ \begin{array}{l}(u,w)\in C_0(\Omega;\mathbb{R}^{d+1}),\\ \big(Hu+p,\,\mathcal{B}w+q\big)(x,y)\in\ |x-y|\,\mathrm{K}\\ \forall\,(x,y)\in G\end{array}\right\}$$

$$=\sup_{\tilde{p},\tilde{q}}\Big\{\langle\tilde{p},\pi\rangle+\langle\tilde{q},\mathscr{S}\rangle\ :\ (\tilde{p},\tilde{q})(x,y)\in\ |x-y|\,\mathrm{K}\quad\forall\,(x,y)\in G\Big\}$$

$$+\sup_{u,w}\Big\{\langle u,f\rangle-\langle Hu,\pi\rangle-\langle\mathcal{B}w,\mathscr{S}\rangle\ :\ (u,w)\in C_0(\Omega;\mathbb{R}^{d+1})\Big\},$$

where, to pass from the first to the second line, we have set $(\tilde{p},\tilde{q}) = (Hu+p, \mathcal{B}w+q)$ which in fact run over all $C(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$ selections of the multifunction $|x-y|\,\mathrm{K}$ whatever is (u,w). This allows to split the supremum into the sum $a(\pi,\mathscr{S})+b(\pi,\mathscr{S})$ where a is the supremum with respect to $(\tilde{p},\tilde{q})$ in the second line and b is the supremum with respect to (u,w) in the third line. By linearity, we see that $b(\pi,\mathscr{S})$ coincides with the indicator of the set of measures $(\pi,\mathscr{S})$ satisfying the conditions (i) and (ii) in (5.89) thus we get $b=\chi_{\Xi(\Omega,f)}$. Next we rewrite $a(\pi,\mathscr{S})$ in the form

$$a(\pi,\mathscr{S}) = \sup\left\{\int \upsilon_1\,d\pi + \int \upsilon_2\,d\mathscr{S} \,:\, \upsilon_1,\upsilon_2\in C(\overline{\Omega}\times\overline{\Omega})\ ,\ (\upsilon_1,\upsilon_2)\in\Gamma(x,y) \text{ on } \overline{\Omega}\times\overline{\Omega}\right\}$$

where $\Gamma(x,y)=|x-y|\,\mathrm{K}$ if $(x,y)\in G$ and $\Gamma(x,y)=\mathbb{R}^2$ if $(x,y)\in\overline{\Omega}\times\overline{\Omega}\setminus G$. Similarly as in the proof of Lemma 5.48 we can check that Γ is lower semi-continuous as a multifunction ranging in the family of closed convex subsets of $\mathbb{R}^2$. According to Theorem 2.23 the support function of $\Gamma(x,y)$ determines a one-homogeneous lower semi-continuous integrand on $\overline{\Omega}^2\times\mathbb{R}^2$: $\chi^*_{\Gamma(x,y)}(s_1,s_2)=|y-x|\,\chi^*_{\mathrm{K}}(s_1,s_2)$ if $(x,y)\in G$ and $\chi^*_{\Gamma(x,y)}(s_1,s_2)=\chi_{\{0,0\}}(s_1,s_2)$ otherwise. Moreover, since $a(\pi,\mathscr{S})$ is the support function of the set of continuous selections of Γ, from Theorem 2.23 the integral representation follows (we use the Goffman-Serrin convention):

$$a(\pi,\mathscr{S}) = \int_{\overline{\Omega}\times\overline{\Omega}} \chi^*_{\Gamma(x,y)}(\pi,\mathscr{S})(dxdy) = \begin{cases} \mathcal{J}(\pi,\mathscr{S}) & \text{if } \mathrm{spt}(\pi,\mathscr{S})\subset G,\\ +\infty & \text{otherwise.}\end{cases}$$

Therefore, recalling that $h^*_G(\pi,\mathscr{S})=a(\pi,\mathscr{S})+b(\pi,\mathscr{S})$, we recover the expression asserted in (i).

Lets us now prove the claim (ii). Convexity of h_G may be showed by exploiting [80, Chapter III, Lemma 2.1] stated for general perturbation functions. Notice that $h_G(0,0)\le 0$ (by taking $(u,w)=(0,0)$). On the other hand, in view of assertion (i), $\Xi_G(\Omega,f)\cap\mathrm{dom}(\mathcal{J})$ is non-empty if and only if $\inf h^*_G<+\infty$. In this case we have $h_G(0,0)\ge h^{**}_G(0,0)=-\inf h^*_G>-\infty$ and the convex function h_G is finite at $(0,0)$ and thus proper.

Assume now that the compact subset G is such that $G\cap\Delta$ is empty. Then there exists $\delta>0$ such that $|y-x|\ge\delta$ for all $(x,y)\in G$. By evaluating the infimum in (5.92) with $(u,w)=(0,0)$ and in view of (5.80), we obtain that

$$h_G(p,q)\le 0 \quad\text{whenever}\quad \max_{(x,y)\in G}\left\{\frac{\big(p(x,y)\big)^2}{2}+q(x,y)\right\}\le\delta,$$

from which follows that the convex function h_G is continuous at $(p,q)=(0,0)$. An important consequence is that h^*_G attains its minimum on $\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$ and that $-h_G(0,0)=-h^{**}_G(0,0)=\min h^*_G$ (see Lemma 2.11), hence the assertion (ii).

We now turn to assertion (iii). Let (p_n,q_n) be a sequence such that $(p_n,q_n)\to 0$ in $C(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)$. By definition of h for each n we may choose $(u_n,w_n)\in C_0(\Omega;\mathbb{R}^{d+1})$ that satisfy (5.94) and furnish $-\int u_n\,df\le h(p_n,q_n)+\frac{1}{n}$ (note that h is proper therefore $h(p_n,q_n)>-\infty$). By virtue of Lemma 5.53 there exists a pair $(u,w)\in\overline{\mathcal{X}}_1$ and a

subsequence u_{n_k} converging to u uniformly on $\overline{\Omega}$. Since (u,w) is admissible for $(\overline{\mathscr{P}})$ there must hold $\int u\,df \leq \mathcal{Z}$. The following chain of inequalities may be written down:

$$\liminf_k h(p_{n_k},q_{n_k}) \geq \liminf_k \left\{-\int u_{n_k}\,df\right\} = -\int u\,df \geq -\mathcal{Z} = h(0,0)$$

where the last equality was given in (5.93). The proof is complete. □

Proof of Theorem 5.50. Acknowleding Proposition 5.14 and equality (5.93) in order to obtain (5.50) we need to show that $-h(0,0) = \inf \widetilde{\mathscr{P}}^*$. Due to assertion (iii) of Lemma 5.54 the function h meets the prerequisite of Lemma 2.11 hence $h(0,0) = h^{**}(0,0)$. Owing to assertion (i) of Lemma 5.54 we find that

$$-h(0,0) = -h^{**}(0,0) = \inf_{(\pi,\mathscr{S})\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}^2)} h^*(\pi,\mathscr{S}) = \inf\Big\{\mathcal{J}(\pi,\mathscr{S}) : (\pi,\mathscr{S})\in\Xi(\Omega,f)\Big\},$$

which proves (5.50).

Let us now take a minimizing sequence $(\pi_n,\mathscr{S}_n)\in\Xi(\Omega,f)$ for $(\widetilde{\mathscr{P}}^*)$. By (5.87), (5.88) the pairs $(\vartheta_{\pi_n},\varsigma_{\mathscr{S}_n})$ are admissible for the problem $(\mathscr{P}^*)$. By acknowledging assertion (iii) of Lemma 5.48 we arrive at

$$\mathcal{Z} = \lim_n \mathcal{J}(\pi_n,\mathscr{S}_n) \geq \limsup_n \int \varrho^0(\vartheta_{\pi_n},\varsigma_{\mathscr{S}_n}) \geq \inf \mathscr{P}^* = \mathcal{Z}$$

therefore $(\vartheta_{\pi_n},\varsigma_{\mathscr{S}_n})$ is a minimizing sequence for $(\mathcal{P}^*)$. □

5.5.2. Constructing an optimal elastic system of strings

In the previous section we have showed that whenever there exists a solution $(\pi,\mathscr{S})\in\mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}\times\mathbb{R}_+)$ of the problem $(\widetilde{\mathscr{P}}^*)$ the pair of fields $(\vartheta_\pi,\varsigma_{\mathscr{S}})\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d\times\mathcal{S}^{d\times d})$ constructed in accordance with (5.85,5.86) solves the problem $(\mathscr{P}^*)$. Consequently, the optimal distribution of Hooke tensor λ in the Optimal Elastic Membrane problem (OEM) in the CanMD setting may be recovered based on $\varsigma_{\mathscr{S}}$ by exploiting Corollary 5.19. Following the spirit of Section 4.8 and Theorem 4.64 we will now show that the step involving computing $\varsigma_{\mathscr{S}}$ is redundant and we can construct $\lambda=\lambda_{\mathscr{S}}$ directly from $\mathscr{S}$.

The case when solution $(\pi,\mathscr{S})$ of $(\widetilde{\mathscr{P}}^*)$ exists is actually the case when optimal elastic membrane λ can be sewn of 1D fibres as sketched in Section 5.2.5 and those fibres become the strings that make up the system $(\pi,\mathscr{S})$. The normalized Hooke field corresponding to a string connecting a pair $(x,y)\in\overline{\Omega}\times\overline{\Omega}\setminus\Delta$ reads

$$\lambda^{x,y} := \tau^{x,y}\otimes\tau^{x,y}\otimes\tau^{x,y}\otimes\tau^{x,y}\;\mathcal{H}^1\llcorner[x,y], \qquad \tau^{x,y} = \frac{y-x}{|y-x|}$$

while we put $\lambda^{x,y}=0$ for $x=y$ by convention. We are about to prove that if $(\pi,\mathscr{S})$ is optimal for $(\widetilde{\mathscr{P}}^*)$ then the Hooke field constructed along the lines of formula (5.37)

$$\langle\Psi,\lambda_{\mathscr{S}}\rangle_{(C,\mathcal{M})} := \frac{2C_0}{\mathcal{Z}}\int_{\overline{\Omega}\times\overline{\Omega}}\left(\int\langle\Psi,\lambda^{x,y}\rangle\right)\mathscr{S}(dxdy) \qquad \forall\,\Psi\in C\big(\overline{\Omega};\mathscr{L}(\mathcal{S}^{d\times d})\big) \tag{5.97}$$

is optimal for the (OEM) problem in the CanMD setting. Beforehand we require an additional feature of solutions $(\pi,\mathscr{S})$:

Proposition 5.55. *Let $(\pi, \mathscr{S})$ be a solution of the problem $(\widetilde{\mathscr{P}^*})$. Then $\pi = \alpha\mathscr{S}$ for a suitable $\alpha \in L^1_{\mathscr{S}}(\overline{\Omega}\times\overline{\Omega})$ and the following rule of equi-repartition of energy holds:*

$$\int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\,\mathscr{S}(dxdy) = \int_{\overline{\Omega}\times\overline{\Omega}} \frac{1}{2}\,|y-x|\,\big(\alpha(x,y)\big)^2\mathscr{S}(dxdy) = \frac{\mathcal{Z}}{2}.$$

Proof. Since $\mathcal{J}(\pi,\mathscr{S}) = \min\widetilde{\mathscr{P}^*} = \mathcal{Z} < \infty$ the fact that $\pi = \alpha\mathscr{S}$ is a restatement of the first part of assertion (ii) in Lemma 5.48. The equi-repartition rule may be showed by exploiting the formula (5.91) and using the fact that $(0,\infty) \ni t \to \mathcal{J}(\alpha\mathscr{S}, t\mathscr{S})$ is minimal at $t = 1$ (note that $(\alpha\mathscr{S}, t\mathscr{S}) \in \Xi(\Omega, f)$ for every $t > 0$). □

For the sake of the proof of the result to come it will be convenient to denote for $(u,w) \in C(\overline{\Omega};\mathbb{R}^{d+1})$ and for $(x,y) \in \overline{\Omega}\times\overline{\Omega}\setminus\Delta$

$$\Lambda(u,w)(x,y) := \frac{1}{2}\,\frac{|u(y)-u(x)|^2}{|y-x|} + \langle w(y)-w(x), \tau^{x,y}\rangle,$$

while for $x = y$ we agree that $\Lambda(u,w)(x,y) = 0$. Clearly a pair of functions $(u,w) \in C_0(\Omega;\mathbb{R}^{d+1})$ is an element of $\overline{\mathcal{X}}_1$ if and only if $\Lambda(u,w)(x,y) \le |y-x|$ for every $(x,y) \in \overline{\Omega}\times\overline{\Omega}$, see (5.79).

Theorem 5.56. *Let $(\pi,\mathscr{S}) = (\alpha\mathscr{S},\mathscr{S})$ be a solution of the problem $(\widetilde{\mathscr{P}^*})$. Then $\lambda_{\mathscr{S}}$ solves the problem of designing a canvas-like membrane* (OEM).

Proof. As the first step we check whether $\lambda_{\mathscr{S}}$ is admissible. First of all, since $\lambda^{x,y} \in \mathcal{M}(\overline{\Omega};\mathcal{K})$ for each (x,y) the measure $\lambda_{\mathscr{S}}$ is an element of $\mathcal{M}(\overline{\Omega};\mathcal{K})$ as well, see assertion (ii) in Theorem 3.35. Then, for Id being the identity operator in $\mathscr{L}(\mathcal{S}^{d\times d})$ we have:

$$\begin{aligned}\int \operatorname{Tr}\lambda_{\mathscr{S}} = \langle \mathrm{Id}, \lambda_{\mathscr{S}}\rangle &= \frac{2C_0}{\mathcal{Z}}\int_{\overline{\Omega}\times\overline{\Omega}}\left(\int_{[x,y]}\langle \mathrm{Id}, \tau^{x,y}\otimes\tau^{x,y}\otimes\tau^{x,y}\otimes\tau^{x,y}\rangle\, d\mathcal{H}^1\right)\mathscr{S}(dxdy)\\ &= \frac{2C_0}{\mathcal{Z}}\int_{\overline{\Omega}\times\overline{\Omega}}|x-y|\,\mathscr{S}(dxdy) = C_0.\end{aligned}$$

where the last equality is due to Proposition 5.55. In order to show the claim it is enough to show that $\mathscr{C}(\lambda_{\mathscr{S}})$ is bounded from above by $\mathscr{C}_{\min}$, which will require a series of inequalities. We start by exploiting the Young's inequality with ε (namely $ab \le \frac{\varepsilon}{2}a^2 + \frac{1}{2\varepsilon}b^2$) in order to obtain:

$$\big(u(y)-u(x)\big)\alpha(x,y) \le \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{3}}\frac{1}{2}\,|x-y|\,\big(\alpha(x,y)\big)^2 + \left(\frac{2C_0}{\mathcal{Z}}\right)^{\frac{1}{3}}\frac{1}{2}\,\frac{|u(y)-u(x)|^2}{|x-y|}. \quad (5.98)$$

Then, since $(\alpha\mathscr{S},\mathscr{S})$ is an element of $\Xi(\Omega,f)$ (see (5.89)) for any $(u,w) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1})$ the virtual work $\int_{\overline{\Omega}} u\,df$ may be written as below and then estimated in the second line by using (5.98):

$$\begin{aligned}\int u\,df &= \int_{\overline{\Omega}\times\overline{\Omega}}\big(u(y)-u(x)\big)\alpha(x,y)\,\mathscr{S}(dxdy) + \left(\frac{2C_0}{\mathcal{Z}}\right)^{\frac{1}{3}}\int_{\overline{\Omega}\times\overline{\Omega}}\langle w(y)-w(x),\tau^{x,y}\rangle\mathscr{S}(dxdy)\\ &\le \left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{3}}\int_{\overline{\Omega}\times\overline{\Omega}}\frac{1}{2}\,|y-x|\big(\alpha(x,y)\big)^2\mathscr{S}(dxdy) + \left(\frac{2C_0}{\mathcal{Z}}\right)^{\frac{1}{3}}\int_{\overline{\Omega}\times\overline{\Omega}}\Lambda(u,w)(x,y)\,\mathscr{S}(dxdy)\\ &\le \frac{1}{2}\left(\frac{\mathcal{Z}^4}{2C_0}\right)^{\frac{1}{3}} + \left(\frac{2C_0}{\mathcal{Z}}\right)^{\frac{1}{3}}\int_{\overline{\Omega}\times\overline{\Omega}}\big(\Lambda(u,w)(x,y)\big)_+\,\mathscr{S}(dxdy)\end{aligned}$$

where to pass to the last line we have again used Proposition 5.55. For any $(u,w) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1})$ the elastic energy enjoys the lower bound:

$$\begin{aligned}\int_{\overline{\Omega}} j_+\big(\lambda_{\mathscr{S}}, \mathcal{A}(u,w)\big) &\geq \int_{\overline{\Omega}\times\overline{\Omega}} \frac{2C_0}{\mathcal{Z}} \left(\int_{\overline{\Omega}} j_+\big(\lambda^{x,y}, \mathcal{A}(u,w)\big)\right) \mathscr{S}(dxdy)\\ &= \int_{\overline{\Omega}\times\overline{\Omega}} \frac{C_0}{\mathcal{Z}} \left(\int_{[x,y]} \Big(\big(\langle \mathcal{A}(u,w), \tau^{x,y}\otimes\tau^{x,y}\rangle\big)_+\Big)^2 d\mathcal{H}^1\right) \mathscr{S}(dxdy)\\ &\geq \int_{\overline{\Omega}\times\overline{\Omega}} \frac{C_0}{|y-x|\,\mathcal{Z}} \left(\int_{[x,y]} \big(\langle \mathcal{A}(u,w), \tau^{x,y}\otimes\tau^{x,y}\rangle\big)_+ d\mathcal{H}^1\right)^2 \mathscr{S}(dxdy)\\ &\geq \int_{\overline{\Omega}\times\overline{\Omega}} \frac{C_0}{|y-x|\mathcal{Z}} \Big(\big(\Lambda(u,w)(x,y)\big)_+\Big)^2 \mathscr{S}(dxdy)\end{aligned}$$

where:

- the first inequality is due to assertion (ii) in Theorem 3.35;
- to pass to the second line we used (5.36);
- to pass to the third line we used Hölder inequality;
- to pass to the last line we used the first inequality in (5.43).

By combining the two chains of inequalities above we can estimate $\mathscr{C}(\lambda_{\mathscr{S}})$ from above:

$$\begin{aligned}&\mathscr{C}(\lambda_{\mathscr{S}}) = \sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{ \int u\, df - \int j\big(\lambda_{\mathscr{S}}, \mathcal{A}(u,w)\big)\right\}\\ &\leq \sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{ \frac{1}{2}\left(\frac{\mathcal{Z}^4}{2C_0}\right)^{\frac{1}{3}} + \int \left(\left(\frac{2C_0}{\mathcal{Z}}\right)^{\frac{1}{3}} \big(\Lambda(u,w)\big)_+ - \frac{C_0}{|y-x|\mathcal{Z}} \Big(\big(\Lambda(u,w)\big)_+\Big)^2\right) d\mathscr{S}\right\}\\ &\leq \frac{1}{2}\left(\frac{\mathcal{Z}^4}{2C_0}\right)^{\frac{1}{3}} + \frac{1}{2}\left(\frac{\mathcal{Z}}{2C_0}\right)^{\frac{1}{3}} \int |y-x|\,\mathscr{S}(dxdy) = \frac{3}{4}\left(\frac{\mathcal{Z}^4}{2C_0}\right)^{\frac{1}{3}} = \mathscr{C}_{\min}\end{aligned}$$

where to pass to the third line we have used the fact that $\sup_{t\in\mathbb{R}}\big\{\big(\frac{2C_0}{\mathcal{Z}}\big)^{\frac{1}{3}} t - \frac{C_0}{|y-x|\mathcal{Z}}\, t^2\big\} = \frac{1}{2}\,|y-x|\,\big(\frac{\mathcal{Z}}{2C_0}\big)^{\frac{1}{3}}$. The proof is complete. □

5.5.3. Optimality conditions for the string systems. Analytical solutions for a one-point-force load

The newly established duality between the displacement-based problem $(\overline{\mathscr{P}})$ and the optimal string system problem $(\widetilde{\mathscr{P}^*})$ opens up a path to optimality conditions for a string system, more precisely for pairs (u,w) and $(\pi,\mathscr{S})$. Since this duality rests upon the two-point duality condition (5.79) we will not be needing additional assumptions on smoothness as it was the case for optimality conditions in Theorem 5.41 where we had to enforce that $(u,w) \in \mathrm{Lip}_0(\Omega;\mathbb{R}^{d+1})$. At this point, however, we will still rely on continuity of both u and w, while from Theorem 5.32 we know that for general solution (u,w) of $(\overline{\mathscr{P}})$ we merely have $w \in BV(\Omega;\mathbb{R}^d)$. Nonetheless, the optimality conditions formulated below will be revisited in Section 5.6.4. Therein, thanks to additional technical results via theory of maximal monotone maps, the optimality criteria will be stated in full generality, i.e. for any $(u,w) \in \overline{\mathcal{X}}_1$.

Proposition 5.57. *Let $(\pi, \mathscr{S})$ be an element of $\mathcal{M}(\overline{\Omega} \times \overline{\Omega}; \mathbb{R} \times \mathbb{R}_+)$ given in the form $(\pi, \mathscr{S}) = (\alpha\mathscr{S}, \mathscr{S})$ with $\alpha \in L^1_{\mathscr{S}}(\overline{\Omega} \times \overline{\Omega})$ and let $(u, w) \in C_0(\Omega; \mathbb{R}^{d+1})$.*

Then the pairs (u, w) and $(\pi, \mathscr{S})$ are optimal for, respectively, $(\overline{\mathscr{P}})$ and $(\widetilde{\mathscr{P}}^*)$ if and only if the following conditions are satisfied:

$$\begin{cases} (i) & (\alpha\mathscr{S}, \mathscr{S}) \in \Xi(\Omega, f), \\ (ii) & \frac{1}{2}\,|u(y) - u(x)|^2 + \langle w(y) - w(x), y - x\rangle \le |y - x|^2 \quad \forall\,(x, y) \in \overline{\Omega} \times \overline{\Omega}, \\ (iii) & \frac{1}{2}\,|u(y) - u(x)|^2 + \langle w(y) - w(x), y - x\rangle = |y - x|^2 \quad \textit{for } \mathscr{S}\textit{-a.e. } (x, y), \\ (iv) & \alpha(x, y) = \frac{u(y) - u(x)}{|y - x|} \qquad \textit{for } \mathscr{S}\textit{-a.e. } (x, y). \end{cases} \tag{5.99}$$

Moreover, if the conditions above are met, then for $\mathscr{S}$-a.e. pair $(x, y) \in \overline{\Omega} \times \overline{\Omega}$ the following statement holds true:

$$u \textit{ and } \langle w, \tau^{x,y}\rangle \textit{ are affine on } [x, y].$$

Remark 5.58. Imposing the form $(\pi, \mathscr{S}) = (\alpha\mathscr{S}, \mathscr{S})$ is not restrictive, cf. Proposition 5.55.

Proof. The admissibility of $(\pi, \mathscr{S}) = (\alpha\mathscr{S}, \mathscr{S})$ in $(\widetilde{\mathscr{P}}^*)$ and of (u, w) in $(\overline{\mathscr{P}})$ are equivalent to conditions (i) and (ii), respectively, see (5.82). Based on (ii) and by exploiting (5.89) we may write

$$\begin{aligned} \int u\,df &= \int_{\overline{\Omega}\times\overline{\Omega}} \left(\alpha(x, y)\big(u(y) - u(x)\big) + \Big\langle w(y) - w(x), \frac{y - x}{|y - x|} \Big\rangle \right) \mathscr{S}(dxdy) \\ &\le \int_{\overline{\Omega}\times\overline{\Omega}} |x - y| \left(1 + \frac{1}{2}\big(\alpha(x, y)\big)^2 \right) \mathscr{S}(dxdy) = \mathcal{J}(\alpha\mathscr{S}, \mathscr{S}), \end{aligned}$$

where to pass to the second line we use the pointwise inequality $(Hu)\,\alpha + \mathcal{B}w \le |y - x|\,\chi^*_{\mathrm{K}}(\alpha, 1)$ (see (5.83)) taking into account that, by condition (ii), the pair $(Hu, \mathcal{B}w)(x, y)$ belongs to $|x - y|\,\mathrm{K}$ for every $(x, y) \in \overline{\Omega} \times \overline{\Omega}$. Therefore, in virtue of the equality $\max \overline{\mathscr{P}} = \inf \widetilde{\mathscr{P}}^*$ asserted in Theorem 5.50, the optimality of $(\pi, \mathscr{S})$ and (u, w) is equivalent to the localized equality:

$$\alpha(x, y)\big(u(y) - u(x)\big) + \Big\langle w(y) - w(x), \frac{y - x}{|y - x|} \Big\rangle = |y - x| \left(1 + \frac{1}{2}\big(\alpha(x, y)\big)^2 \right)$$

that must hold for $\mathscr{S}$-a.e. (x, y). After multiplying by $|y - x|$, the equality can be rewritten as

$$\frac{1}{2}\,|u(y) - u(x)|^2 + \langle w(y) - w(x), y - x\rangle - |y - x|^2 = \frac{1}{2}\Big(u(y) - u(x) - |y - x|\,\alpha(x, y)\Big)^2.$$

From (ii), we see that the latter equality holds if and only if (iii) and (iv) are both satisfied. The "moreover" part of the theorem is a direct consequence of optimality condition (iii) owing to Proposition 5.37. □

Remark 5.59. The newly derived optimality conditions may be utilized to show that solution of $(\widetilde{\mathscr{P}}^*)$ fails to exist in general – it is enough to point to an example of data Ω, f where the system (5.99) is infeasible. Let us revisit the radial-symmetric Example 5.45 where the uniform load $f = p\,\mathcal{L}^2 \llcorner \Omega$ furnishes a strictly increasing function $\mathcal{F}$. Then the

formula (5.75) provides a Lipschitz continuous solution (u, w) of the problem $(\overline{\mathscr{P}})$ such that u is strictly concave. In particular, for any pair $(x, y) \in \overline{\Omega} \times \overline{\Omega} \setminus \Delta$ the function u is not affine on $[x, y]$. Thus, owing to the "moreover part" of Proposition 5.57, we conclude that the system (5.99) cannot be solved for any string system $(\pi, \mathscr{S})$. This way we get a *counter-example to the existence of a string system* for the optimal membrane problem.

We shall next employ optimality conditions in Proposition 5.57 to solve the optimal membrane problem for a one-point-force load $f = \delta_{x_0}$ for $x_0 \in \Omega$ (we choose a normalized, unit force for simplicity) and a wide range of convex domains Ω. It will appear that optimal membrane is a system of strings that connect the point x_0 directly to the boundary $\partial\Omega$. In the first place we will tackle the case of disk domain Ω. This result will be then utilized to give solutions for other convex domains (but not arbitrary, cf. Remark 5.61).

One-point-force load in a disk

Let $\Omega \subset \mathbb{R}^2$ be a disk of radius R and centred at the origin. We consider a load $f = \delta_{x_0}$ with $x_0 \in \Omega$. It is convenient to introduce the geometric parameter $d_0 := \sqrt{R^2 - |x_0|^2}$. Let $p \in \mathrm{P}(\partial\Omega)$ be any probability measure satisfying the barycentre condition

$$x_0 = [p] := \int_{\partial\Omega} x\, p(dx). \tag{5.100}$$

Then we consider the pair $(\pi, \mathscr{S}) = (\alpha\mathscr{S}, \mathscr{S}) \in \big(\mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})\big)^2$ defined by

$$\pi = p(dx) \otimes \delta_{x_0}(dy), \quad \mathscr{S} = \frac{|x_0 - x|}{\sqrt{2}\, d_0}\, p(dx) \otimes \delta_{x_0}(dy), \quad \alpha(x, x_0) = \frac{\sqrt{2}\, d_0}{|x - x_0|} \quad \text{for } x \in \partial\Omega. \tag{5.101}$$

We claim that such a pair solves the problem $(\widetilde{\mathscr{P}^*})$, thus providing a truss solution $(\vartheta_\pi, \varsigma_\mathscr{S})$ to $(\mathscr{P}^*)$ in virtue of Theorem 5.50. First we show that $(\pi, \mathscr{S})$ is admissible for $(\widetilde{\mathscr{P}^*})$. Indeed $\mathscr{S} \geq 0$ and condition (i) in (5.89) is fulfilled since $p(\partial\Omega) = 1$ while condition (ii) follows from (5.100). In view of Proposition 5.57, it remains to find an admissible pair $(u, w) \in C_0(\Omega; \mathbb{R}^{d+1})$ satisfying the conditions $(ii), (iii), (iv)$ in (5.99). To that aim, we proceed in polar coordinates (r, θ) with respect to x_0 so that the domain Ω is parametrized as

$$\Omega := \Big\{ x_0 + r\, \mathrm{e}_r(\theta) \ : \ \theta \in [0, 2\pi), \quad 0 \leq r < \rho(\theta) \Big\}\,,$$

where $\mathrm{e}_r(\theta) := (\cos\theta, \sin\theta)$ and $\rho(\theta) : [0, 2\pi) \to [R - |x_0|, R + |x_0|]$ is a Lipschitz function. Then we propose the following u, w constructed from the function h whose graph coincides with the conical surface in $\mathbb{R}^3$ with $(x_0, 1)$ as its vertex and containing $\partial\Omega \times \{0\}$, namely:

$$(u, w) = \Big(\sqrt{2}\, d_0, 2\, x_0\Big)\, h \qquad \text{where} \quad h(x_0 + r\, \mathrm{e}_r) := 1 - \frac{r}{\rho(\theta)}, \tag{5.102}$$

see Fig. 5.4(b). Since (u, w) vanishes on the whole boundary $\partial\Omega$ and $(u, w)(x_0) = (\sqrt{2}\, d_0, 2\, x_0)$, the equalities in (iii) and (iv) in (5.99) become, respectively

$$d_0^2 + \langle 2x_0, x_0 - x \rangle = |x_0 - x|^2, \qquad \alpha(x, x_0) = \frac{u(x_0)}{|x - x_0|} = \frac{\sqrt{2}\, d_0}{|x - x_0|}.$$

The second equality then follows from the explicit form of $\alpha(x, x_0)$ in (5.101) while the first equality, equivalent to $d_0^2 + |x_0|^2 = |x|^2$, holds at every $x \in \partial\Omega$ hence p-a.e (note

that $|x| = R$ on $\partial\Omega$). Eventually it remains to check the two-point condition (ii) being the admissibility condition: $(u, w) \in \overline{\mathcal{X}}_1$. As (u, w) is Lipschitz, this condition may be validated by checking the equivalent point-wise condition in Proposition 5.38, namely that at every point of differentiability $x \in \Omega$ the eigenvalues of the symmetric matrix $A(x) := \frac{1}{2}\nabla u(x) \otimes \nabla u(x) + e(w)(x)$ do not exceed 1. We proceed with computations using the frame $(\mathrm{e}_r, \mathrm{e}_\theta)$ (where $\mathrm{e}_\theta(\theta) = (-\sin\theta, \cos\theta)$). In view of (5.102) we have for $x = x_0 + r\,\mathrm{e}_r$:

$$\nabla u = \sqrt{2}\, d_0\, \nabla h \;=\; \sqrt{2}\, d_0\, (a\,\mathrm{e}_r + a'\mathrm{e}_\theta), \qquad e(w) = 2\, x_0 \overset{s}{\otimes} \nabla h \qquad \text{where} \quad a(\theta) := -\frac{1}{\rho(\theta)}.$$

We obtain the decomposition $A(x) = A_{1,1}\,\mathrm{e}_r \otimes \mathrm{e}_r + 2\,A_{1,2}\,\mathrm{e}_r \overset{s}{\otimes} \mathrm{e}_\theta + A_{2,2}\,\mathrm{e}_\theta \otimes \mathrm{e}_\theta$ in which

$$A_{1,1} = d_0^2\, a^2 + 2\langle x_0, \mathrm{e}_r\rangle\, a, \quad A_{1,2} = d_0^2\, aa' + \langle x_0, \mathrm{e}_r\rangle\, a' + \langle x_0, \mathrm{e}_\theta\rangle\, a,$$
$$A_{2,2} = d_0^2\, (a')^2 + 2\langle x_0, \mathrm{e}_\theta\rangle\, a'. \tag{5.103}$$

Fortunately the latter expressions simplify. Indeed, since $|x_0 + \rho\,\mathrm{e}_r|^2 = R^2$, we have $\rho^2 + 2\rho\langle x_0, \mathrm{e}_r\rangle = R^2 - |x_0|^2 = d_0^2$ and therefore:

$$d_0^2\, a^2 = 1 - 2\langle x_0, \mathrm{e}_r\rangle\, a, \qquad d_0^2\, aa' = -\langle x_0, \mathrm{e}_r\rangle\, a' - \langle x_0, \mathrm{e}_\theta\rangle\, a,$$

where the second equality is obtained by differentiating the first one with respect to θ. Thus $A_{1,1} = 1$ and $A_{1,2} = 0$. Next we prove that $A_{2,2} \le 0$. To that aim we exploit the fact that $u = 0$ on $\partial\Omega$ so that ∇u is parallel to $x_0 + \rho\,\mathrm{e}_r$. This implies more relations:

$$(\langle x_0, \mathrm{e}_r\rangle + \rho)\, a' = \langle x_0, \mathrm{e}_\theta\rangle\, a, \qquad d_0^2\, (a')^2 = \frac{\langle x_0, \mathrm{e}_\theta\rangle^2}{(\langle x_0, \mathrm{e}_r\rangle + \rho)^2}\big(1 - 2\langle x_0, \mathrm{e}_r\rangle\, a\big).$$

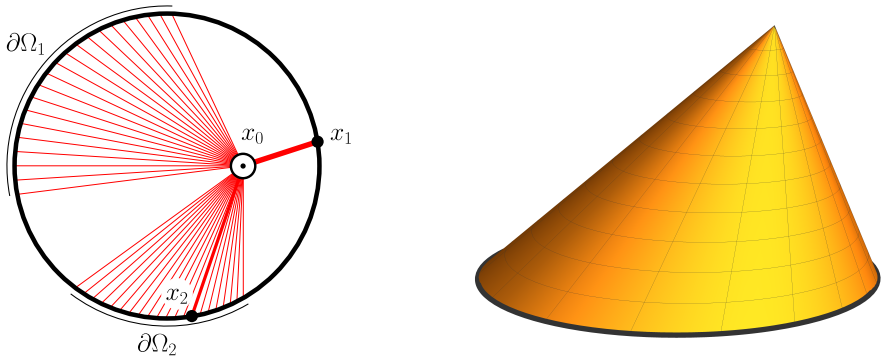
Substituting in the expression for $A_{2,2}$ in (5.103), we get:

$$A_{2,2} = \frac{\langle x_0, \mathrm{e}_\theta\rangle^2}{(\langle x_0, \mathrm{e}_r\rangle + \rho)^2}\Big(1 - 2\langle x_0, \mathrm{e}_r\rangle\, a + 2(\langle x_0, \mathrm{e}_r\rangle + \rho)\, a\Big) = -\frac{\langle x_0, \mathrm{e}_\theta\rangle^2}{(\langle x_0, \mathrm{e}_r\rangle + \rho)^2}.$$

Summarizing we have proved than $A = A(x)$ is a diagonal matrix with eigenvalues not larger that 1. This implies the admissibility condition (ii) and the optimality of (u, w) for the dual problem while $(\pi, \mathscr{S})$ given by (5.101) is optimal for $(\widetilde{\mathscr{P}^*})$. Moreover the minimal energy is given by

$$\mathscr{Z} = \min \widetilde{\mathscr{P}^*} = \min \mathscr{P}^* = \max \overline{\mathscr{P}} = \sqrt{2}\, d_0.$$

Condition (5.100) imposed on the probability measure $p \in \mathcal{M}_+(\partial\Omega)$ is very mild, offering a wide choice of string system solutions $(\pi, \mathscr{S})$ via formulas (5.101). For instance we may consider $p = \sum_{i=1}^2 b_i\, \mathcal{H}^1 \llcorner \partial\Omega_i + \sum_{i=1}^2 B_i\, \delta_{x_i}$, with $b_i, B_i \in R_+$ chosen so that $p(\partial\Omega) = 1$ and $x_0 = \int_{\partial\Omega} x\, p(dx)$, then the field $\varsigma_{\mathscr{S}}$ in accordance with (5.86) is visualized in Fig. 5.4(a). This optimal pre-stress field $\varsigma_{\mathscr{S}}$ may be divided into three parts: a "continuous" fan of strings radiating from x_0 and supported on the arc $\partial\Omega_1$; a 1D string $[x_1, x_0]$ of finite cross section area; the hybrid part consisting both of continuous fan supported on $\partial\Omega_2$ and a string $[x_2, x_0]$.



(a) (b)

Figure 5.4: Optimal membrane problem for a disk domain Ω and a one-point-force load $f = \delta_{x_0}$: (a) pre-stress field $\varsigma_{\mathscr{S}}$ optimal for $(\mathscr{P}^*)$; (b) the conic graph of function u solving $(\overline{\mathscr{P}})$.

One-point-force load in a general convex domain

The construction used above for the disk paves a way for finding a solution for other shapes of design domains. More precisely let $\Omega \subset \mathbb{R}^2$ be a general bounded convex domain and let $x_0 \in \Omega$. We introduce the condition:

$$\text{there exists} \quad y_0 \in \Omega \quad \text{such that} \quad x_0 \in \operatorname{co}\big(\Pi^{\partial\Omega}(y_0)\big) \tag{5.104}$$

where $\Pi_{\partial\Omega}(y_0)$ is the minimal set of $d(y_0, \cdot)$ on $\partial\Omega$, see (4.74). Then we define

$$d_0(y_0) := \sqrt{\big(d^{\partial\Omega}(y_0)\big)^2 - |x_0 - y_0|^2}.$$

Proposition 5.60. *Assume that* (5.104) *is satisfied for a suitable* $y_0 \in \Omega$ *and let* p *be any probability supported on* $\Pi^{\partial\Omega}(y_0)$ *such that* $x_0 = \int x\, p(dx)$. *Then the pair* $(\pi, \mathscr{S})$ *given by*

$$\pi = p(dx) \otimes \delta_{x_0}(dy), \qquad \mathscr{S} = \frac{|x_0 - x|}{\sqrt{2}\, d_0(y_0)}\, p(dx) \otimes \delta_{x_0}(dy) \tag{5.105}$$

solves the problem $(\widetilde{\mathscr{P}}^*)$ *for* $f = \delta_{x_0}$ *and we have* $\min \widetilde{\mathscr{P}}^* = \min \mathscr{P}^* = \sqrt{2}\, d_0(y_0)$.

Proof. In the same way as for the disk, the equalities $p(\partial\Omega) = 1$ and $x_0 = \int x\, p(dx)$ imply that $(\pi, \mathscr{S})$ given in (5.105) is an admissible pair for $(\widetilde{\mathscr{P}}^*)$. Next we observe that (5.104) implies that $x_0 \in B(y_0, R)$ where $R = d(y_0, \partial\Omega)$. Without loss of generality we may assume that y_0 is the origin so that $d_0(y_0) = \sqrt{R^2 - |x_0|^2}$. Therefore $(\pi, \mathscr{S})$ given in (5.105) coincides with the optimal pair we found for the disk $B(0, R)$ (see (5.101)). Let us consider $(\bar{u}, \bar{w})$ being the zero extension to Ω of the pair (u, w) given by (5.102) in the disk $B(0, R)$. Clearly $(\bar{u}, \bar{w})$ is Lipschitz continuous, vanishes on the whole boundary $\partial\Omega$ and satisfies the pointwise gradient constraint (5.61) in the whole Ω. Therefore the condition (ii) in the system (5.99) is fulfilled on $\overline{\Omega} \times \overline{\Omega}$. On the other hand, the conditions (iii) and (iv) are obviously satisfied in the same way as for the disk. That way we recover the optimality of $(\pi, \mathscr{S})$ for the problem $(\widetilde{\mathscr{P}}^*)$ set on the domain Ω. The corresponding minimal energy is given by $\min \widetilde{\mathscr{P}}^* = \sqrt{2}\,\sqrt{R^2 - |x_0|^2} = \sqrt{2}\, d_0(y_0)$. □

Remark 5.61. Due to the condition (5.104) which is crucial in our previous construction, we can only handle single loads $f = \delta_{x_0}$ when x_0 belongs to the compact subset $K_0 := \bigcup \{\mathrm{co}(\Pi_{\partial\Omega}(y_0)) : y_0 \in \overline{\Omega}\}$ which is determined by those points y_0 for which $\Pi_{\partial\Omega}(y_0)$ has more than one element. In contrast with the square for which $K_0 = \overline{\Omega}$, it happens that K_0 is a strict subset of $\overline{\Omega}$ for an ellipse of large excentricity.

Example 5.62. As an illustration of Proposition 5.60 we can solve the one-force problem for a rectangular design domain Ω of sides $2R$ and $2R + L$ where $L > 0$ as presented in Fig. 5.5 where also locations of strategic points $a_1, \ldots, a_6$ are specified. We choose $f = \delta_{x_0}$ for $x_0 \in \Omega$. The dashed lines in Fig 5.5 partition the domain into seven regions. For the

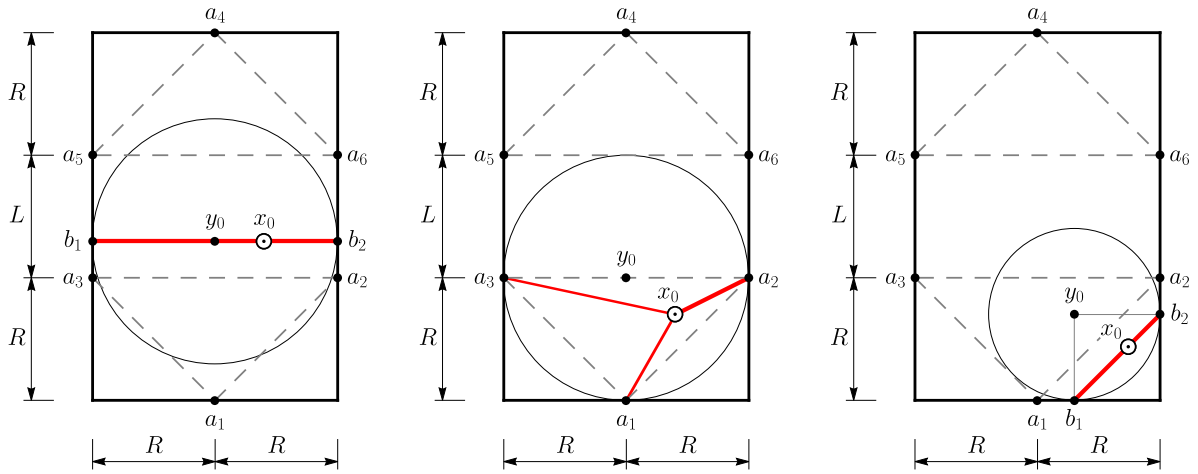


Figure 5.5: Location of y_0 in accordance with Proposition 5.60 and optimal $\varsigma_{\mathscr{S}}$ for three positions of x_0

position x_0 lying inside each of those regions, the form of solution $(\pi, \mathscr{S})$ given by (5.105) differs. By symmetry it is not restrictive to assume that x_0 belongs to the south-east part of Ω so that we are reduced to three cases. For each of them we precise the selected y_0 and the probability p. The corresponding optimal ς is depicted in Fig 5.5.

Case (a): x_0 *belongs to the interior of* $\mathrm{co}\{a_2, a_3, a_5, a_6\}$. By b_1 and b_2 we denote the orthogonal projections of x_0 onto the longer sides of the domain, see Fig. 5.5(a). The unique y_0 satisfying (5.104) is given by $y_0 = (b_1 + b_2)/2$ (thus $\Pi_{\partial\Omega}(y_0) = \{b_1, b_2\}$). The unique element p of $\mathrm{P}(\Pi^{\partial\Omega}(y_0))$ such that $x_0 = \int x\, p(dx)$ is given by $p = \frac{|x_0 - b_2|}{|b_1 - b_2|}\delta_{b_1} + \frac{|x_0 - b_1|}{|b_1 - b_2|}\delta_{b_2}$. After some simplifications, we arrive at a minimal energy given by $\min \widetilde{\mathscr{P}}^* = \sqrt{2\,|x_0 - b_1||x_0 - b_2|}$. The optimal pair $(\vartheta_\pi, \varsigma_{\mathscr{S}})$ represented in Fig. 5.5(a) corresponds to a one-dimensional string $[b_1, b_2]$.

Case (b): x_0 *belongs to the closure of* $\mathrm{co}\{a_1, a_2, a_3\}$. The unique y_0 satisfying (5.104) is given by $y_0 = (a_2 + a_3)/2$. Then $\Pi^{\partial\Omega}(y_0) = \{a_1, a_2, a_3\}$ and p is the unique probability $p = \sum_{i=1}^3 \alpha_i\, \delta_{a_i}$ with $\alpha_i \geq 0$ such that $x_0 = \int x\, p(dx) = \sum_{i=1}^3 \alpha_i\, a_i$. The optimal membrane $\varsigma_{\mathscr{S}}$ consists of three strings tied at x_0 as depicted in Fig. 5.5(b).

Case (c): x_0 *lies in* $\mathrm{co}\{a_1, a_2, a_7\}$. The unique possible y_0 is determined by its projections b_1, b_2 on the boundary where $b_1 \in [a_1, a_7]$ and $b_2 \in [a_2, a_7]$ are chosen so that the segment $[b_1, b_2]$ contains x_0 and is parallel to $[a_1, a_2]$ (see Fig. 5.5(c)). Then $\Pi^{\partial\Omega}(y_0) = \{b_1, b_2\}$ and $p = \frac{|x_0 - b_2|}{|b_1 - b_2|}\delta_{b_1} + \frac{|x_0 - b_1|}{|b_1 - b_2|}\delta_{b_2}$. Similarly to the Case (a), the optimal membrane is a one dimensional string $[b_1, b_2]$.

In addition we can handle the case of a square where $L = 0$, $a_3 = a_5$ and $a_2 = a_6$. For x_0 lying in the triangle $a_1a_2a_7$, the construction of y_0 and p are the same as in the Case (c) above rendering the one dimensional string $[b_1, b_2]$ the unique solution. In contrast, if x_0 lies in the interior of the rotated square $\mathrm{co}\{a_1, a_2, a_3, a_4\}$, the unique possible y_0 is the center of the square while $\Pi^{\partial\Omega}(y_0) = \{a_1, a_2, a_3, a_4\}$. Then there are infinitely many probability measures $p = \sum_{i=1}^4 \alpha_i \, \delta_{a_i}$ that give $x_0 = \sum_{i=1}^4 \alpha_i \, a_i$. Among the corresponding solutions, we recover in particular the four-string structure that was advertised in the introduction, see Fig. 5.1(b). △

5.5.4. The conic quadratic program as discretization of the pair $(\overline{\mathscr{P}})$, $(\widetilde{\mathscr{P}^*})$

By developing the duality theory that employs the two-point condition we have laid the foundation for an efficient numerical method tailored for the optimal membrane problem in the setting of Canvas-like Material Design setting (CanMD). Once again we will exploit the mathematical similarities of this particular setting of (OEM) problem to the Michell problem for which the ground structure numerical method have been effectively used for over a half of century, cf. [75].

In this section we continue to assume that Ω is convex. Thanks to the detailed duality analysis in Section 5.5.1 formulating the ground structure approximation of the optimal membrane problem is straightforward: in the problem $(\widetilde{\mathscr{P}^*})$ we are looking for those string systems $(\pi, \mathscr{S})$ whose nodes are contained in a pre-fixed finite set $\mathrm{X}^h \subset \overline{\Omega}$, namely we enforce that $\mathrm{spt}\,(\pi, \mathscr{S}) \subset \mathrm{X}^h \times \mathrm{X}^h$. We have prepared for such an occasion in Lemma 5.54: we choose $G = \mathrm{X}^h \times \mathrm{X}^h$ and the approximated optimal string problem that we are looking for is $(\widetilde{\mathscr{P}^*_G})$, which further will be shortly denoted by $(\widetilde{\mathscr{P}^*_h})$. In general we will be needing to work with a discretization

$$f_h \in \mathcal{M}(\mathrm{X}^h; \mathbb{R})$$

of the load f; naturally we should perform the discretization so that $f_h \overset{*}{\rightharpoonup} f$ when the grid parameter h goes to zero (we will define f_h explicitly later). Let us first write down the ground structure approximation of the primal problem $(\mathscr{P})$ which is simply $-h_G(0,0) = -h_{\mathrm{X}^h \times \mathrm{X}^h}(0,0)$, recall (5.92) for definition of the perturbation function h (not to be confused with the grid parameter):

$$\mathcal{Z}_h := \sup \left\{ \int u \, df_h \;\middle|\; \begin{array}{l} (u, w) \in C_0(\Omega; \mathbb{R}^{d+1}), \\ \big(Hu, \mathcal{B}w\big)(x, y) \in |y - x| \, \mathrm{K} \qquad \forall\, (x, y) \in \mathrm{X}^h \times \mathrm{X}^h \end{array} \right\} \qquad (\mathscr{P}_h)$$

which defines the approximate value function $\mathcal{Z}_h$. We can always define $G = \mathrm{X}^h \times \mathrm{X}^h \setminus \Delta$ since excluding the diagonal does not affect h_G. Owing to finiteness of X^h such set G remains compact and by virtue of statement (ii) of Lemma 5.54 we have zero-gap result:

$$\mathcal{Z}_h = \inf\Big\{\mathcal{J}(\pi,\mathscr{S}) \ : \ (\pi,\mathscr{S}) \in \Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega, f_h)\Big\} \qquad (\widetilde{\mathscr{P}}_h^*)$$

(it can be showed that $\sup \mathscr{P}_h = \inf \widetilde{\mathscr{P}}_h^* = \infty$ whenever $\Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega, f_h) \cap \mathrm{dom}(\mathcal{J}) = \varnothing$).

The need for discretizing the load can be clarified as follows: condition $(\pi,\mathscr{S}) \in \Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega, f_h)$ implies that the equilibrium equations (see (5.89)) of a string system must be satisfied in a strict sense and not approximatively as it was the case for the set $\Sigma_h(\Omega, f)$ defined for FEM in Section 4.6. On top of that all pairs $(\pi,\mathscr{S})$ in the set $\Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega, f_h)$ must be supported in $\mathrm{X}^h \times \mathrm{X}^h$. This way we learn that $\Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega, f_h)$ is an empty set once $\mathrm{spt}\, f_h \nsubseteq \mathrm{X}^h$. In Proposition 5.65 we will show that the condition $\mathrm{spt}\, f_h \subset \mathrm{X}^h$ is sufficient for non-emptiness of $\Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega, f_h)$ if we make an additional assumption on the set X^h itself:

$$\mathrm{X}^h \backslash \partial\Omega \subset \mathrm{int}\big(\mathrm{co}(\mathrm{X}^h \cap \partial\Omega)\big). \tag{5.106}$$

Meanwhile a simple lower bound for $\mathcal{Z}_h$ may be given based solely on the inclusion $\Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega, f) \subset \Xi(\Omega, f)$ for any f:

Proposition 5.63. *For a given grid* X^h *assume that* $\mathrm{spt}\, f \subset \mathrm{X}^h$ *and put* $f_h = f$. *Then the inequality as follows hold true:*

$$\mathcal{Z} \le \mathcal{Z}_h.$$

In the sequel we will borrow a few notions introduced in Section 4.6 on the finite element method for the pair $(\overline{\mathcal{P}}), (\mathcal{P}^*)$, starting with the triangulation $\mathcal{T}^h = \{E_i^h\}$ that, although is not intrinsic to the ground structure method, will be convenient to use. The numerical method will be developed for the 2D ambient space only (the most meaningful one in terms of mechanics) and for simplicity the domain $\Omega \subset \mathbb{R}^2$ will be assumed to be polygonal. We introduce a finite dimensional space of Lipschitz continuous functions (u, w) that vanish on the boundary and are affine on each element:

$$\mathcal{X}^h(\Omega) := \Big\{(u,w) \in C(\overline{\Omega};\mathbb{R}^{d+1}) \ : \ (u,w) = 0 \text{ on } \partial\Omega, \\ \big(u(x), w(x)\big) = a_i + A_i x \ \ \forall\, x \in E_i^h, \ a_i \in \mathbb{R}^{d+1}, \ A_i \in \mathbb{R}^{(d+1)\times d}\Big\}.$$

For X^h we shall take the finite set of all vertices of triangulation $\mathcal{T}^h$. We will henceforward assume that $\partial\Omega$ is the union of a subfamily of edges (treated as closed segments) of the triangulation. In that case, with $\mathrm{X}_0^h := \mathrm{X}^h \cap \partial\Omega$, the space $\mathcal{X}^h(\Omega)$ is isomorphic to the $3m$-dimensional space of nodal displacements at vertices $\mathrm{X}^h \setminus \mathrm{X}_0^h$, where $m = \mathrm{card}(\mathrm{X}^h \setminus \mathrm{X}_0^h)$. We introduce an ordering of the vertices: $\{v_1, \ldots, v_m\} = \mathrm{X}^h \setminus \mathrm{X}_0^h$ and a unique interpolation operator $\iota^h = (\iota_\perp^h, \iota_{\shortparallel}^h) : \mathbb{R}^{m\times 3} \to \mathcal{X}^h(\Omega)$ being the isomorphism that for $\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^m$ satisfies

$$\begin{aligned}(u,w) \ &= \iota^h(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2) \\ &= \big(\iota_\perp^h(\mathbf{u}), \iota_{\shortparallel}^h(\mathbf{w}_1, \mathbf{w}_2)\big)\end{aligned} \quad \Leftrightarrow \quad \begin{cases} u(v_j) = \mathbf{u}(j) \\ w(v_j) = \mathbf{w}_1(j)\,\mathrm{e}_1 + \mathbf{w}_2(j)\,\mathrm{e}_2 \end{cases} \quad \forall\, j \in \{1, \ldots, m\}.$$

where $\mathrm{e}_1, \mathrm{e}_2$ is a fixed Cartesian basis. We define the load vector $\mathbf{f} \in \mathbb{R}^m$ which is uniquely determined through the linear mapping

$$\langle \mathbf{u}, \mathbf{f} \rangle = \int_{\overline{\Omega}} \iota_{\perp}^h(\mathbf{u})\, df \qquad \forall\, \mathbf{u} \in \mathbb{R}^m. \tag{5.107}$$

We may now propose an explicit formula for discrete load $f_h \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ by putting

$$f_h := \sum_{j=1}^{m} \mathbf{f}(j)\, \delta_{v_j} \quad \in \quad \mathcal{M}(\mathrm{X}^h; \mathbb{R}). \tag{5.108}$$

It is straightforward to show that $f_h \overset{*}{\rightharpoonup} f$ when the mesh parameter h tends to zero.

The ground structure is essentially the finite system of strings identified with segments interconnecting the nodes in X^h. The number of strings in the ground structure equals $N = \big(\mathrm{card}(\mathrm{X}^h) - 1\big)\, \mathrm{card}(\mathrm{X}^h)/2$. We introduce functions $\mathrm{x}, \mathrm{y} : \{1, \dots, N\} \to \mathbb{R}^2$ such that the map

$$[\mathrm{x}(\,\cdot\,), \mathrm{y}(\,\cdot\,)] : \{1, \dots, N\} \;\to\; \Big\{ [x, y] \,:\, (x, y) \in \mathrm{X}^h \times \mathrm{X}^h \setminus \Delta \Big\}$$

is a bijection, i.e. $\mathrm{x}(k)$ and $\mathrm{y}(k)$ are the starting and, respectively, the ending point of the k-th bar in the ground structure. Effectively we thus impose orientation of each string. We define the vector $\mathbf{L} \in \mathbb{R}^N$ of the lengths of the strings, namely

$$\mathbf{L}(k) := \big| \mathrm{y}(k) - \mathrm{x}(k) \big|.$$

When restricted to the finite ground structure the linear geometric operators $H, \mathcal{B}$ may be written in the form of matrices $\mathbf{H}, \mathbf{B}_1, \mathbf{B}_2 \in \mathbb{R}^{N \times m}$ that are uniquely determined by linear mappings:

$$\big(\mathbf{H}\mathbf{u}\big)(k) = \Big(H\big(\iota_{\perp}^h(\mathbf{u})\big) \Big) \Big(\mathrm{x}(k), \mathrm{y}(k) \Big), \tag{5.109}$$

$$\big(\mathbf{B}_1\mathbf{w}_1 + \mathbf{B}_2\mathbf{w}_2\big)(k) = \Big(\mathcal{B}\big(\iota_{\shortparallel}^h(\mathbf{w}_1, \mathbf{w}_2)\big) \Big) \Big(\mathrm{x}(k), \mathrm{y}(k) \Big). \tag{5.110}$$

Remark 5.64. The three matrices $\mathbf{H}, \mathbf{B}_1, \mathbf{B}_2$ do not depend on the particular operators $\iota_{\perp}^h$, $\iota_{\shortparallel}^h$ that make the piece-wise affine interpolations. Indeed, the right hand sides of (5.109), (5.110) depend on values of $\iota_{\perp}^h(\mathbf{u})$ and $\iota_{\shortparallel}^h(\mathbf{w}_1, \mathbf{w}_2)$ at vertices $v_j \in \mathrm{X}^h$ only and those are determined solely by $\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2$. In the end the interpolation operators are used for the sake of convenience only.

For a given pair $(u, w) \in C_0(\Omega; \mathbb{R}^{d+1})$ we may rewrite the admissibility condition from $(\mathscr{P}_h)$ in the matrix/vector format:

$$\frac{1}{2} \frac{\big((\mathbf{H}\mathbf{u})(k)\big)^2}{\mathbf{L}(k)} + (\mathbf{B}_1\mathbf{w}_1 + \mathbf{B}_2\mathbf{w}_2)(k) \le \mathbf{L}(k) \qquad \text{for every } k \tag{5.111}$$

where $\mathbf{u}(j) = u(v_j)$, $\mathbf{w}_1(j) = \langle w(v_j), \mathrm{e}_1 \rangle$, $\mathbf{w}_2(j) = \langle w(v_j), \mathrm{e}_2 \rangle$. Using the extra condition (5.106) on the grid X^h we can provide a discrete variant of the compactness result in Theorem 5.25:

Proposition 5.65. *Assume that the nodal gird* X^h *satisfies* (5.106). *Then the set*

$$\Big\{ (\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2) \in \mathbb{R}^{3m} \,:\, (5.111) \text{ holds true} \Big\} \tag{5.112}$$

is convex and compact. Consequently $\mathcal{Z}_h < \infty$ *and* $\Xi_{\mathrm{X}^h \times \mathrm{X}^h}(\Omega, f_h) \neq \varnothing$.

Proof. The set $\bigl\{(t_1,t_2)\in\mathbb{R}^2 : \frac{1}{2}\,(t_1)^2/\mathbf{L}(k)+t_2\le\mathbf{L}(k)\bigr\}$ is convex and closed for each k which implies convexity and closedness of the set (5.112).

Showing the boundedness of (5.112) is more involved. We start by noticing that thanks to assumption (5.106) the Hausdorff distance $\delta := D_H\bigl(\mathrm{X}^h\setminus\partial\Omega\,,\,\partial\bigl(\mathrm{co}(\mathrm{X}^h\cap\partial\Omega)\bigr)\bigr)$ is positive by finiteness of X^h. Then for arbitrary $x\in\mathrm{X}^h\setminus\partial\Omega$ and arbitrary $\eta\in S^{d-1}$ there exists $y=y_{x,\eta}\in\mathrm{X}^h\cap\partial\Omega$ such that $\langle\eta,x-y_{x,\eta}\rangle\ge\delta$. Indeed, the point $y_{x,\eta}$ may always be chosen such that the half-space $H=\bigl\{z\in\mathbb{R}^d : \langle z-y_{x,\eta},\eta\rangle\ge 0\bigr\}$ contains $\mathrm{co}(\mathrm{X}^h\cap\partial\Omega)$ and therefore contains $B(x,\delta)$ as well. Naturally, if η is not of unit length we have $\langle\eta,x-y_{x,\eta}\rangle\ge\delta\,|\eta|$.

For fixed vectors $(\mathbf{u},\mathbf{w}_1,\mathbf{w}_2)$ we set $(u,w)=\iota^h(\mathbf{u},\mathbf{w}_1,\mathbf{w}_2)$. Let us choose $x=v_j$ for $j\in\{1,\dots,m\}$ and $\eta=\mathbf{w}_1(j)\,\mathrm{e}_1+\mathbf{w}_2(j)\,\mathrm{e}_2$. We know that $w(x)=\eta$ while $w(y_{x,\eta})=0$. From the condition $\bigl(Hu,\mathcal{B}w\bigr)(y_{x,\eta},x)\in|x-y_{x,\eta}|\,\mathrm{K}$ we infer that

$$|x-y_{x,\eta}|^2\ge\langle w(x)-w(y_{x,\eta}),x-y_{x,\eta}\rangle=\langle\eta,x-y_{x,\eta}\rangle\ge\delta\,|\eta|=\delta\sqrt{\bigl(\mathbf{w}_1(j)\bigr)^2+\bigl(\mathbf{w}_2(j)\bigr)^2}.$$

Since the left hand side in the chain above is not grater than $\bigl(\mathrm{diam}(\Omega)\bigr)^2$ and since $\delta>0$ is independent of $(\mathbf{w}_1,\mathbf{w}_2)$ we obtain uniform upper bound for the norm of the latter pair in $\mathbb{R}^{2m}$. Now we can exploit the full form of the condition $\bigl(Hu,\mathcal{B}w\bigr)(y_{x,\eta},x)\in|x-y_{x,\eta}|\,\mathrm{K}$ to infer that $\mathbf{u}(j)=u(x)$ is bounded as well.

The finiteness of $\mathcal{Z}_h=\sup\mathscr{P}_h$ is straightforward: for $(u,w)\in C_0(\Omega;\mathbb{R}^{d+1})$ by definition (5.108) we have $\int u\,df_h=\langle\mathbf{u},\mathbf{f}\rangle$ for $\mathbf{u},\mathbf{w}_1,\mathbf{w}_2$ defined below (5.111) and the claim follows by the first part of this proposition. Since $\inf\mathscr{P}_h^*=\mathcal{Z}_h<\infty$ the non-emptiness of $\Xi_{\mathrm{X}^h\times\mathrm{X}^h}(\Omega,f_h)$ follows. □

In the process of proving Proposition 5.65 we have showed that problem $(\mathscr{P}_h)$ gets to be rewritten as a convex finite dimensional programming problem where with variables $\mathbf{u}$, $\mathbf{w}_1$, $\mathbf{w}_2\in\mathbb{R}^m$ we maximize the linear form $\langle\mathbf{u},\mathbf{f}\rangle$ under the constraint (5.111) that must hold for each string k. The constraint itself is convex quadratic, which distinguishes this ground structure method from the one used for Michell problem, see Section 4.7.3. In case of truss approximation of Michell structures one ends up with a pair of mutually dual linear programming problems $(\mathbf{MP}_{\mathrm{X}^h})$, $(\mathbf{MP}^*_{\mathrm{X}^h})$. Nontheless it is well established (see e.g. [9]) that programs with linear objective functional and convex quadratic constraint may be reformulated as a conic quadratic programming problem in accordance with Section 2.6. To that aim we shall employ the so called rotated quadratic cone in $\mathbb{R}^3$:

$$\mathcal{K}^3_{\mathrm{rot,quad}}=\biggl\{\mathbf{z}\in\mathbb{R}^3\ :\ \mathbf{z}(1),\mathbf{z}(2)\ge 0,\quad 2\,\mathbf{z}(1)\,\mathbf{z}(2)\ge\bigl(\mathbf{z}(3)\bigr)^2\biggr\}.$$

Recall that $\mathcal{K}^3_{\mathrm{rot,quad}}$ is self-dual, cf. (2.30). The finite dimensional conic quadratic reformulation of problem $(\mathscr{P}_h)$ can be readily given:

$$\mathcal{Z}_h=\sup_{\substack{\mathbf{u},\,\mathbf{w}_1,\,\mathbf{w}_2\in\mathbb{R}^m\\ \mathbf{t}_1,\mathbf{t}_2\in\mathbb{R}^m_+,\ \mathbf{t}_3\in\mathbb{R}^m}}\left\{\langle\mathbf{u},\mathbf{f}\rangle\ :\ \begin{array}{c}\mathbf{t}_1=\mathbf{L}\\ \mathbf{t}_2+\mathbf{B}_1\mathbf{w}_1+\mathbf{B}_2\mathbf{w}_2=\mathbf{L}\\ \mathbf{t}_3+\mathbf{H}\mathbf{u}=\mathbf{0}\\ \bigl(\mathbf{t}_1(k),\mathbf{t}_2(k),\mathbf{t}_3(k)\bigr)\in\mathcal{K}^3_{\mathrm{rot,quad}}\qquad\forall\,k\end{array}\right\}\qquad(\boldsymbol{P}_h)$$

One can easily check that the auxiliary variables $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ can be eliminated so that the quadratic constraint (5.111) is recast. Below we shall prove that the dual conic quadratic program reads:

$$\inf_{\mathbf{r},\mathbf{s}\in\mathbb{R}^m_+,\ \mathbf{p}\in\mathbb{R}^m} \left\{ \langle \mathbf{L}, \mathbf{r}\rangle + \langle \mathbf{L}, \mathbf{s}\rangle \ : \begin{array}{c} \mathbf{H}^\top \mathbf{p} = \mathbf{f} \\ \mathbf{B}_1^\top \mathbf{s} = \mathbf{0} \\ \mathbf{B}_2^\top \mathbf{s} = \mathbf{0} \\ \big(\mathbf{r}(k), \mathbf{s}(k), \mathbf{p}(k)\big) \in \mathcal{K}^3_{\mathrm{rot,quad}} \qquad \forall\, k \end{array} \right\} \tag{$\boldsymbol{P}_h^*$}$$

By eliminating the auxiliary variable $\mathbf{r}$ one could show that problem $(\boldsymbol{P}_h^*)$ is a finite dimensional reformulation of problem $(\widetilde{\mathscr{P}}_h^*)$. Consequently, duality between $(\boldsymbol{P}_h)$ and $(\boldsymbol{P}_h^*)$ is, in a way, already established owing to the duality result between $(\mathscr{P}_h)$ and $(\widetilde{\mathscr{P}}_h^*)$ that was deduced from Lemma 5.54. Notwithstanding this, we readdress the duality problem for the pair above to adhere to the classical algorithm for conic programming problems recalled in Section 2.6. In the process an algebraic counterpart of the system of optimality conditions (5.99) will be given:

Theorem 5.66. *Let $\mathrm{X}^h \subset \overline{\Omega}$ be a finite grid that satisfies condition* (5.106)*. Then $(\boldsymbol{P}_h), (\boldsymbol{P}_h^*)$ is a pair of mutually dual conic quadratic problems such that*

$$\mathcal{Z}_h = \max \boldsymbol{P}_h = \min \boldsymbol{P}_h^* < \infty,$$

in particular both problems admit their solutions.

Moreover $(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2) \in \mathbb{R}^{3\times m}$ and $(\mathbf{r}, \mathbf{s}, \mathbf{p}) \in \mathbb{R}^N_+ \times \mathbb{R}^N \times \mathbb{R}^N_+$ solve problems $(\boldsymbol{P}_h)$ and, respectively, $(\boldsymbol{P}_h^*)$ if and only if the optimality conditions below hold true:

$$\left\{ \begin{array}{lll} (i) & \mathbf{H}^\top \mathbf{p} = \mathbf{f}, \quad \mathbf{B}_1^\top \mathbf{s} = \mathbf{0}, \quad \mathbf{B}_2^\top \mathbf{s} = \mathbf{0}, & \\ (ii) & \frac{1}{2}\frac{((\mathbf{Hu})(k))^2}{\mathbf{L}(k)} + (\mathbf{B}_1\mathbf{w}_1 + \mathbf{B}_2\mathbf{w}_2)(k) \le \mathbf{L}(k) & \text{for every } k, \\ (iii) & \frac{1}{2}\frac{((\mathbf{Hu})(k))^2}{\mathbf{L}(k)} + (\mathbf{B}_1\mathbf{w}_1 + \mathbf{B}_2\mathbf{w}_2)(k) = \mathbf{L}(k) & \text{if } \ \mathbf{s}(k) \neq 0, \\ (iv) & \mathbf{p}(k) = \frac{(\mathbf{Hu})(k)}{\mathbf{L}(k)}\,\mathbf{s}(k) \qquad \text{for every } k, & \\ (v) & \mathbf{r}(k) = \frac{1}{2}\frac{(\mathbf{p}(k))^2}{\mathbf{s}(k)} \quad \text{if } \ \mathbf{s}(k) \neq 0 \qquad \text{and} \qquad \mathbf{r}(k) = 0 \quad \text{otherwise.} & \end{array} \right. \tag{5.113}$$

Proof. We prove the result by putting the pair of problems $(\boldsymbol{P}_h)$ and $(\boldsymbol{P}_h^*)$ in the framework of the canonical conic programming problem presented in Section 2.6: the problem $(\boldsymbol{P}_h^*)$ becomes the primal problem (CP) given therein, while $(\boldsymbol{P}_h)$ the dual problem (CP*) provided that we set:

$$\mathbf{x} = \begin{bmatrix} \mathbf{r} \\ \mathbf{s} \\ \mathbf{p} \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} \mathbf{L} \\ \mathbf{L} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{H}^\top \\ \mathbf{0} & \mathbf{B}_1^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_2^\top & \mathbf{0} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} \mathbf{u} \\ \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix}, \quad \boldsymbol{\eta}_k = \begin{bmatrix} \mathbf{t}_1(k) \\ \mathbf{t}_2(k) \\ \mathbf{t}_3(k) \end{bmatrix};$$

$\mathbf{Q}_k \in \mathbb{R}^{3\times 3N}$ for $k \in \{1, \dots, N\}$ are 0 and 1 matrices with non-zero entries as follows: $\mathbf{Q}_k\big(1, k\big) = \mathbf{Q}_k(2, 2k) = \mathbf{Q}_k(3, 3k) = 1$; for each k we choose the cone $\mathcal{K}_i = \mathcal{K}^3_{\mathrm{rot,quad}}$. The problem $(\boldsymbol{P}_h^*)$ is thus recast by (CP) directly and, since $\mathcal{K}^3_{\mathrm{rot,quad}}$ is self-dual (see the end of Section 2.6), the problem (CP*) becomes exactly $(\boldsymbol{P}_h)$ as well.

We now show that solution of $(\boldsymbol{P}_h)$ exists and $\max \boldsymbol{P}_h < \infty$. The auxiliary variables $\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3$ may be eliminated thus returning to the quadratic constraint (5.111). Then, owing to Proposition 5.65 the set of admissible vectors $(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2)$ is compact in $\mathbb{R}^{3\times m}$ thus furnishing the claim.

Next we check the assumptions $(1), (2)$ in Definition 2.70. It is easy to verify that $\bigcap_{k=1}^{N} \operatorname{Ker} \mathbf{Q}_k = \{0\} \subset \mathbb{R}^{3N}$ from which (2) follows. Linear independence of rows of $\mathbf{A}$ is equivalent to the fact that the kernel of the linear operator $(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2) \mapsto \big(\mathbf{Hu}, \mathbf{B}_1\mathbf{w}_2 + \mathbf{B}_2\mathbf{w}_2\big) \in \mathbb{R}^N \times \mathbb{R}^N$ is the trivial space $\{\mathbf{0}\} \subset \mathbb{R}^{3m}$. It must be the case since otherwise Proposition 5.65 would be false. By virtue of statement (i) in Theorem 2.71 equality $\max \boldsymbol{P}_h = \inf \boldsymbol{P}_h^*$ holds and solution to $(\boldsymbol{P}_h)$ exists. By assertion (iii) of Theorem 2.71 in order to show existence of solution to $(\boldsymbol{P}_h^*)$ we must show strict feasibility (see Definition 2.70) of $(\mathrm{CP}^*) = (\boldsymbol{P}_h)$. This part is easy, since $\mathbf{u} = \mathbf{w}_1 = \mathbf{w}_2 = \mathbf{0}$ and $\mathbf{t}_1 = \mathbf{L}$, $\mathbf{t}_2 = \mathbf{L}$, $\mathbf{t}_3 = \mathbf{0}$ is in fact a strictly feasible solution to $(\boldsymbol{P}_h)$.

To prove the "moreover part" we exploit assertion (iv) in Theorem 2.71: the vectors $\mathbf{x}$ and $(\mathbf{y}, \boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_N)$ that are feasible for (CP) and (CP^*), respectively, solve those problem if and only if the complementary slackness condition $\langle\, \mathbf{Q}_k\mathbf{x}\,,\, \boldsymbol{\eta}_k \,\rangle = 0$ holds for each k, i.e. here $\mathbf{r}(k)\,\mathbf{t}_1(k) + \mathbf{s}(k)\,\mathbf{t}_2(k) + \mathbf{p}(k)\,\mathbf{t}_3(k) = 0$. Feasibility in $(\boldsymbol{P}_h)$ is equivalent to satisfying condition (ii) in (5.113) and moreover to the equalities $\mathbf{t}_1(k) = \mathbf{L}(k)$, $\mathbf{t}_2(k) = \mathbf{L}(k) - (\mathbf{B}_1\mathbf{w}_1 + \mathbf{B}_2\mathbf{w}_2)(k)$, $\mathbf{t}_3(k) = (\mathbf{Hu})(k)$; these conditions, along with (i) in (5.113) will thus be assumed to be true in the remainder of the proof. Once $\mathbf{s}(k) = 0$ the conic constraint in $(\boldsymbol{P}_h^*)$ implies that $\mathbf{p}(k) = 0$ while the complementary slackness condition gives $\mathbf{r}(k) = 0$. If $\mathbf{s}(k) > 0$ then the complementary slackness condition may be rewritten as:

$$\begin{aligned}
&\left(\frac{1}{2}\frac{\big((\mathbf{Hu})(k)\big)^2}{\mathbf{L}(k)} + (\mathbf{B}_1\mathbf{w}_1 + \mathbf{B}_2\mathbf{w}_2)(k)\right)\mathbf{s}(k)\\
&= \mathbf{L}(k)\,\mathbf{s}(k) + \frac{1}{2}\,\mathbf{L}(k)\,\mathbf{s}(k)\left(\frac{(\mathbf{Hu})(k)}{\mathbf{L}(k)} - \frac{\mathbf{p}(k)}{\mathbf{s}(k)}\right)^2 + \mathbf{L}(k)\left(\mathbf{r}(k) - \frac{1}{2}\frac{\big(\mathbf{p}(k)\big)^2}{\mathbf{s}(k)}\right).
\end{aligned}$$

Due to (ii) in (5.113) the LHS is not greater than $\mathbf{L}(k)\,\mathbf{s}(k)$ whereas the RHS is not smaller than $\mathbf{L}(k)\,\mathbf{s}(k)$ owing to the conic constraint $\big(\mathbf{r}(k), \mathbf{s}(k), \mathbf{p}(k)\big) \in \mathcal{K}^3_{\mathrm{rot,quad}}$. Ultimately the equality holds if and only if conditions (iii), (iv), (v) in (5.113) are fulfilled. □

With the use of optimality conditions (5.113) from solutions of the pair $(\boldsymbol{P}_h)$ and $(\boldsymbol{P}_h^*)$ we can construct solutions of problems $(\mathscr{P}_h)$ and $(\widetilde{\mathscr{P}}_h^*)$ respectively:

Proposition 5.67. *Assume that the triples* $(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2) \in \mathbb{R}^{3\times m}$ *and* $(\mathbf{r}, \mathbf{s}, \mathbf{p}) \in \mathbb{R}^N_+ \times \mathbb{R}^N \times \mathbb{R}^N_+$ *solve problems* $(\boldsymbol{P}_h)$ *and* $(\boldsymbol{P}_h^*)$ *respectively. Then the pair*

$$u_h = \iota^h_{\perp}(\mathbf{u}), \qquad w_h = \iota^h_{\shortparallel}(\mathbf{w}_1, \mathbf{w}_2)$$

solve problem $(\mathscr{P}_h)$ *while the pair*

$$\pi_h = \sum_{k=1}^{N} \mathbf{p}(k)\,\delta_{(\mathrm{x}(k),\mathrm{y}(k))}, \qquad \mathscr{S}_h = \sum_{k=1}^{N} \mathbf{s}(k)\,\delta_{(\mathrm{x}(k),\mathrm{y}(k))}$$

solves $(\widetilde{\mathscr{P}}_h^*)$.

Proof. The fact that (u_h, w_h) is a solution is due to the already established equivalence between problems $(\mathscr{P}_h)$ and $(\boldsymbol{P}_h)$.

The support of the pair $(\pi_h, \mathscr{S}_h)$ is clearly contained in $\mathrm{X}^h \times \mathrm{X}^h$ therefore to verify the admissibility condition $(\pi_h, \mathscr{S}_h) \in \Xi_{\mathrm{X}^h \times \mathrm{X}^h}(\Omega, f_h)$ we must check if equations $H^* \pi_h = f_h$, $\mathcal{B}^* \mathscr{S}_h = 0$ hold true. Since $\operatorname{spt} f_h \subset \mathrm{X}^h$ it may be checked that those equations reduce to $\mathbf{H}^\top \mathbf{p} = \mathbf{f}$, $\mathbf{B}_1^\top \mathbf{s} = \mathbf{0}$, $\mathbf{B}_2^\top \mathbf{s} = \mathbf{0}$ and feasibility of $(\pi_h, \mathscr{S}_h)$ follows. Using optimality condition (iv) in (5.113) we see that $\pi_h \ll \mathscr{S}_h$ and therefore formula (5.91) may be used to compute

$$\mathcal{J}(\pi_h, \mathscr{S}_h) = \sum_{k=1}^{N} \mathbf{L}(k) \left(1 + \frac{1}{2}\Big(\frac{\mathbf{p}(k)}{\mathbf{s}(k)}\Big)^2\right) \mathbf{s}(k) = \sum_{k=1}^{N} \Big(\mathbf{L}(k)\, \mathbf{s}(k) + \mathbf{L}(k)\, \mathbf{r}(k)\Big)$$

where in the last equality we have exploited the optimality condition (v) in (5.113). The right hand side above is exactly the value function in $(\boldsymbol{P}_h^*)$ therefore it is equal to $\mathcal{Z}_h$ by Theorem 5.66, which renders $(\pi_h, \mathscr{S}_h)$ a solution of $(\widetilde{\mathscr{P}}_h^*)$. □

Similarly as for the finite element method in Section 4.6 the actual convergence analysis of the numerical ground structure method put forth above lies beyond of the scope of this thesis. In fact, according to the author's knowledge, convergence has not been examined for the original ground structure method for Michell structures, although plenty of numerical experiments compare very favourably to analytical solutions, cf. [118]. Both topics are of great interest for the present author and are scheduled for his future research. Ideally, one can hope for the following convergence result:

Conjecture 5.68. *For X^h being a family of nodal grids that satisfy condition* (5.106) *let (u_h, w_h) and $(\pi_h, \mathscr{S}_h)$ be constructed in accordance with Proposition 5.67. Then,*

$$\mathcal{Z}_h \to \mathcal{Z}$$

when the grid parameter h goes to zero. Moreover:

(i) there exists a pair $(u, w) \in \overline{\mathcal{X}}_1 \subset C_0(\Omega) \times (BV \cap L^\infty)(\Omega; \mathbb{R}^2)$ solving $(\overline{\mathscr{P}})$ such that (possibly upon choosing a subsequence) for h approaching zero

$$u_h \to u \quad \text{uniformly in } \Omega, \qquad w_h \to w \quad \text{in } L^1(\Omega; \mathbb{R}^2);$$

(ii) there exists a pair $(\vartheta, \varsigma) \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^2 \times \mathcal{S}^{2\times 2})$ solving $(\mathscr{P}^*)$ such that (possibly upon choosing a subsequence) for h approaching zero

$$(\vartheta_{\pi_h}, \varsigma_{\mathscr{S}_h}) \overset{*}{\rightharpoonup} (\vartheta, \varsigma) \quad \text{in } \mathcal{M}(\overline{\Omega}; \mathbb{R}^2 \times \mathcal{S}_+^{2\times 2})$$

where $\vartheta_{\pi_h}, \varsigma_{\mathscr{S}_h}$ are constructed via (5.85), (5.86), *respectively.*

Following the footsteps of Section 5.5.2 from solution $\mathbf{s} \in \mathbb{R}^N$ of problem $(\boldsymbol{P}_h^*)$ one can directly construct a Hooke field $\lambda_h = \lambda_{\mathscr{S}_h} \in \mathcal{M}(\overline{\Omega}; \mathcal{K}_{fib})$ that is supported on the ground structure (being a family of segments) and that ought to approximate a solution λ of the (OEM) problem in the CanMD setting:

$$\lambda_{\mathscr{S}_h} = \frac{2C_0}{\mathcal{Z}_h} \sum_{k=1}^{N} \mathbf{s}(k)\, \lambda^{\mathrm{x}(k), \mathrm{y}(k)}. \tag{5.114}$$

If we define the approximate compliance $\mathscr{C}_h = \mathscr{C}_h(\lambda)$ by replacing f with f_h in the original definition (5.9), then one can use almost identical arguments as in the proof of Theorem 5.56 to show that $\int \operatorname{Tr} \lambda_{\mathscr{S}_h} = C_0$ and

$$\mathscr{C}_h(\lambda_{\mathscr{S}_h}) \;\leq\; \frac{3}{4}\left(\frac{\mathcal{Z}_h^4}{2C_0}\right)^{\frac{1}{3}}.$$

Accordingly, we make another conjecture about convergence of the numerical method:

Conjecture 5.69. *For* X^h *being a family of nodal grids satisfying* (5.106) *let* $(\mathbf{r},\mathbf{s},\mathbf{p}) \in \mathbb{R}^N_+ \times \mathbb{R}^N \times \mathbb{R}^N_+$ *solve problem* $(\boldsymbol{P}_h^*)$*. Then there exists a solution* $\lambda \in \mathcal{M}(\overline{\Omega};\mathcal{K}_{fib})$ *that solves the* (OEM) *problem in CanMD setting such that (possibly upon choosing a subsequence)*

$$\lambda_{\mathscr{S}_h} \overset{*}{\rightharpoonup} \lambda \quad \text{in } \mathcal{M}(\overline{\Omega};\mathcal{K}_{fib})$$

when h *goes to zero.*

5.5.5. Numerical implementation and the adaptive "member adding" approach

In the numerical simulations we shall choose polygonal domains Ω that are mainly convex (an adaptation of the ground stricture method towards non-convex domains will be mentioned in Example 5.75). For a chosen resolution parameter $h > 0$ we will use regular grids of the form

$$\mathrm{X}^h = \Big(\big\{(h\,j_1, h\,j_2) \,:\, j_1, j_2 \in \mathbb{Z}\big\} \cap \Omega\Big) \cup \mathrm{X}^h_{\partial\Omega} \tag{5.115}$$

where $\mathrm{X}^h_{\partial\Omega}$ is a union of sets of the form $\big(\{h\,j_1\}\times\mathbb{R}\big)\cap\partial\Omega$ and $\big(\mathbb{R}\times\{h\,j_2\}\big)\cap\partial\Omega,\ j_1, j_2 \in \mathbb{Z}$, without their relative interiors (horizontal and vertical lines may intersect with whole edges of polygon Ω). Condition (5.106) may be easily verified for X^h as above. The regular grid (5.115) is not fit for handling complicated geometries of the design domains – for that purpose the GRAND method developed in [172] could be employed.

Similarly to the FE method developed in Section 4.6 the ground-structure-based numerical method for the optimal membrane problem was implemented in MATLAB® R2018b. After building the vectors $\mathbf{L} \in \mathbb{R}^N, \mathbf{f} \in \mathbb{R}^m$ and matrices $\mathbf{H}, \mathbf{B}_1, \mathbf{B}_2$ (in a sparse form) the conic quadratic programs $(\boldsymbol{P}_h)$, $(\boldsymbol{P}_h^*)$, precisely in the form written in the previous subsection, is being solved via dedicated solver from the MOSEK® 8 toolbox, see [10]. As a result one obtains solutions $(\mathbf{r},\mathbf{s},\mathbf{p})$ and $(\mathbf{u}_1,\mathbf{u}_2,\mathbf{w},\mathbf{t}_1,\mathbf{t}_2,\mathbf{t}_3)$ of the both programs. The post-processing part, including mainly visualizations, is performed in Mathematica® 11.1.

During the tests run on a personal computer, the MOSEK solver was capable of solving the conic program $(\boldsymbol{P}_h)$, $(\boldsymbol{P}_h^*)$ for ground structures containing up to several million strings, which for a square domain Ω roughly corresponds to a 50×50 grid X^h of the form (5.115) – for a convex domain the number of strings is equal to $N = \bar{m}\,(\bar{m}-1)/2$ where $\bar{m}$ is the total number of points in X^h. For comparison, a 200×200 grid generates $m \approx 800\cdot10^6$ bars, being far beyond the reach of any interior-point method solver being run on a PC. Very similar computational challenges are faced when solving the large scale linear programming problems $(\mathbf{MP}_{\mathrm{X}^h})$, $(\mathbf{MP}^*_{\mathrm{X}^h})$ in order to produce accurate truss approximations of Michell structures. In [91] the "rectangular structure" of the $N \times m$ matrices $\mathbf{B}_1, \mathbf{B}_2$ ($N >> m$ for large m) was taken advantage of. The authors proposed an adaptive approach where as

the point of departure one takes a small ground structure consisting of $\widetilde{N} \approx 4\,\bar{m}$ members connecting only the neighbouring nodes in X^h; after solving the linear program for matrices $\widetilde{\mathbf{B}}_1, \widetilde{\mathbf{B}}_2$ the ground structure is then updated in subsequent iterations by adding bars that violate the feasibility condition in the problem $(\mathbf{MP}^*_{\mathrm{X}^h})$. The so called *member-adding* adaptive technique was further developed in [158] and also successfully applied to other linear programming problems: slip-line analysis in [93], optimal design of grillages in [32], optimal design of long-span bridges with gravity loading in [84]. In this thesis we bring the member-adding adaptive method to the pair of conic quadratic programs $(\boldsymbol{P}_h)$, $(\boldsymbol{P}^*_h)$. Below we outline the algorithm:

{1} Preparing the first iteration:

{1.1} Throughout the algorithm the nodal grid X^h according to (5.115) is fixed for a chosen resolution parameter h. For the number $m = \#(\mathrm{X}^h \backslash \partial\Omega)$ the load vector $\mathbf{f} \in \mathbb{R}^m$ is built and does not change in subsequent iterations. The *full ground structure* will be the family of all N members/segments interconnecting the nodes in X^h that are contained in the set $\overline{\Omega}$ (each index $k \in \{1, \ldots, N\}$ identifies the segment $[\mathrm{x}(k), \mathrm{y}(k)]$).

{1.2} For the first iteration $iter = 1$ the initial *active ground structure* is prepared that consists only of members connecting the neighbouring nodes in X^h, namely the set of member indices k is restricted to

$$I_1 = \Big\{ k \in \{1, \ldots, N\} \,:\, |\mathrm{x}(k) - \mathrm{y}(k)|_\infty \leq h \Big\},$$

where $|x|_\infty = \max\{|x_1|, |x_2|\}$ for $x = (x_1, x_2) \in \mathbb{R}^2$. For $N_1 = \#(I_1)$ vector $\mathbf{L}^{(1)} \in \mathbb{R}^{N_1}$ and matrices $\mathbf{H}^{(1)}, \mathbf{B}^{(1)}_1, \mathbf{B}^{(1)}_2 \in \mathbb{R}^{N_1 \times m}$ are built (in their sparse form) according to formulas (5.109), (5.110) yet for rows referring only to indices $k \in I_1$.

{2} The adaptive loop starting at the counter $iter = 1$. Each iteration $iter$ involves the following steps:

{2.1} The MOSEK solver is run for the conic quadratic program $(\boldsymbol{P}^{(iter)}_h), (\boldsymbol{P}^{*,(iter)}_h)$ that is constructed for vectors and matrices $\mathbf{f} \in \mathbb{R}^m$, $\mathbf{L}^{(iter)} \in \mathbb{R}^{N_{iter}}$, $\mathbf{H}^{(iter)}, \mathbf{B}^{(iter)}_1, \mathbf{B}^{(iter)}_2 \in \mathbb{R}^{N_{iter} \times m}$. A solution is returned:

$$\mathbf{u}^{(iter)}, \mathbf{w}^{(iter)}_1, \mathbf{w}^{(iter)}_2 \in \mathbb{R}^m, \qquad \mathbf{r}^{(iter)}, \mathbf{s}^{(iter)}, \mathbf{p}^{(iter)} \in \mathbb{R}^{N_{iter}}.$$

{2.2} The set of members in the full ground structure, for which the two-point condition is violated by $\mathbf{u}^{(iter)}, \mathbf{w}^{(iter)}_1, \mathbf{w}^{(iter)}_2$ being competitors in the original problem $(\boldsymbol{P}_h)$, is identified:

$$\delta I_{iter} = \Big\{ k \in \{1, \ldots, N\} \,:\, \frac{1}{2} \frac{\big((\mathbf{H}\mathbf{u}^{(iter)})(k)\big)^2}{\mathbf{L}(k)} + (\mathbf{B}_1 \mathbf{w}^{(iter)}_1 + \mathbf{B}_2 \mathbf{w}^{(iter)}_2)(k) > \mathbf{L}(k) \Big\};$$

it must be stressed that all the members k in the *full* ground structure are checked.

{2.3} If $\delta I_{iter} = \varnothing$ then the loop is terminated. Otherwise the active ground structure is updated:

$$I_{iter+1} = I_{iter} \cup \delta I_{iter}$$

and the vectors/matrices are appended to $\mathbf{L}^{(iter+1)} \in \mathbb{R}^{N_{iter+1}}$, $\mathbf{H}^{(iter+1)}, \mathbf{B}_1^{(iter+1)}, \mathbf{B}_2^{(iter+1)} \in \mathbb{R}^{N_{iter+1} \times m}$ by adding rows corresponding to every $k \in \delta I_{iter}$, where $N_{iter+1} = \#(I_{iter+1})$. Next the step {2.1} is repeated for $iter := iter + 1$.

{3} Constructing solutions of the original conic quadratic program $(\boldsymbol{P}_h)$, $(\boldsymbol{P}_h^*)$:
For $iter$ being the last counter in the adaptive loop we set $\big(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2\big) = \big(\mathbf{u}^{(iter)}, \mathbf{w}_1^{(iter)}, \mathbf{w}_2^{(iter)}\big)$ while $\mathbf{r}, \mathbf{s}, \mathbf{p} \in \mathbb{R}^N$ are constructed by allocating components of $\mathbf{r}^{(iter)}, \mathbf{s}^{(iter)}, \mathbf{p}^{(iter)} \in \mathbb{R}^{N_{iter}}$ into zero vectors in $\mathbb{R}^N$. Optimality conditions (5.113) are then met for the triples $\big(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2\big)$ and $\big(\mathbf{r}, \mathbf{s}, \mathbf{p}\big)$: conditions $(i), (iii), (iv), (v)$ are numerically established by MOSEK solver at each iteration, while condition (ii) holding for every $k \in \{1, \ldots, N\}$ (i.e. for each member in the full ground structure) is guaranteed by $\delta I_{iter} = \varnothing$.

A few comments on the algorithm are in order. First, we must ensure that in the first iteration $iter = 1$ the problems $(\boldsymbol{P}_h^{(1)}), (\boldsymbol{P}_h^{*,(1)})$ have solutions. It is important to note that verification of condition (5.106) is *a priori* insufficient to answer affirmatively, since the set I_1 could potentially miss crucial members that guaranteed existence of feasible solutions in problems $(\boldsymbol{P}_h)$, $(\boldsymbol{P}_h^*)$ posed for the full ground structure. However, the proof of Proposition 5.65 may be slightly modified to furnish a suitable compactness result in the first iteration. Indeed, for each point $\bar{x} = (\bar{x}_1, \bar{x}_2) \in \mathrm{X}^h \cap \Omega$ it is enough to exploit the two point conditions for strings of lengths h parallel to e_1 or e_2 going from $\bar{x}$ to $(\bar{x}_1 - a_1, \bar{x}_2), (\bar{x}_1 + b_1, \bar{x}_2), (\bar{x}_1, \bar{x}_2 - a_2), (\bar{x}_1, \bar{x}_2 + b_2) \in \mathrm{X}^h_{\partial\Omega}$ for some $a_1, a_2, b_1, b_2 > 0$. Ultimately assertions of Theorem 5.66 hold true for problems $(\boldsymbol{P}_h^{(1)}), (\boldsymbol{P}_h^{*,(1)})$ and thus also for the problems in the subsequent iterations.

The adaptive loop is by construction guaranteed to finish in a finite number of steps that does not exceed $N - N_1$ which, of course, is very pessimistic. Although no rate of convergence is examined in this thesis, the member-adding algorithm usually converges in no more than 10 to 15 iterations while (for large N of magnitudes: 10^6 to 10^9) the number N_{iter} is usually kept on the level between $N/1000$ and $N/100$. Exceptions may occur if in the solution of the "continuous" problem $(\overline{\mathscr{P}})$ we have subregions of Ω where $\frac{1}{2}\nabla u \otimes \nabla u + e(w)$ equals to identity matrix I exactly – then we may expect that every single string lying in this subregion will be added to the set of active bars I_{iter} which makes N_{iter} to blow up with each iteration. This phenomena is known from truss and grillage optimization and can be handled by a simple trick: from each component of the vector $\mathbf{L}^{(iter)}$ we subtract a very small constant, e.g. 10^{-7}, that is independent of string index k.

5.5.6. Numerical simulations

For different domains Ω (mainly for a square domain) and loads f we will present numerical solutions of the optimal membrane problem obtained via the developed ground

structure method. Using the "member adding" adaptive algorithm implemented in MATLAB the pair of conic quadratic problems $(\boldsymbol{P}_h)$, $(\boldsymbol{P}_h^*)$ are solved. Computations are performed on a mobile workstation with Intel Core i7-6700HQ processor, 16 GB RAM and running 64-bit Windows 7. Each problem is solved for three different resolution parameters h. The computational details are summarized in Table 5.1 whose entries are described in detail in the first Example 5.70 for which additional analytical computations will be provided. It should be stressed that symmetry of the investigated problems has not been exploited numerically. In each example we will display the optimal string system in the form of a 1D measure $\varsigma_{\mathscr{S}_h} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}_+^{2\times 2})$ (see Proposition 5.67) and, preferably, $\vartheta_{\pi_h} \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^2)$, $u_h \in C_0(\Omega)$, $w_h \in C_0(\Omega; \mathbb{R}^2)$.

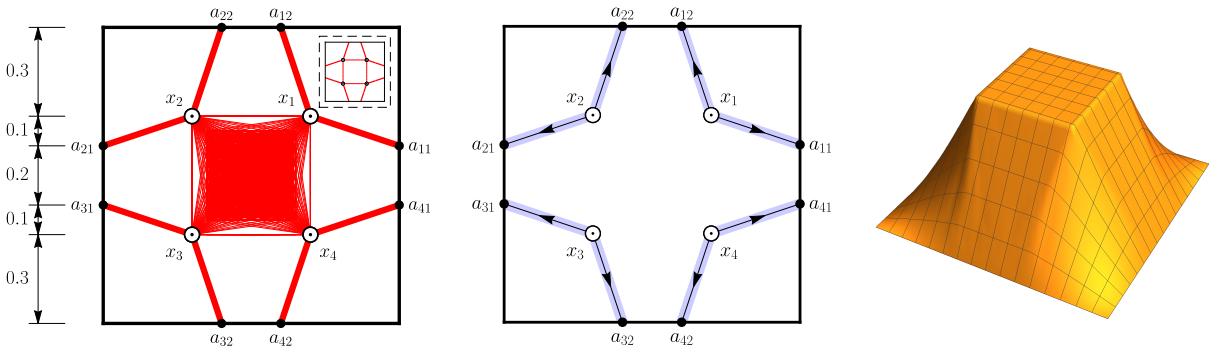
Table 5.1: Summary on numerical computations for the pair of problems $(\boldsymbol{P}_h)$, $(\boldsymbol{P}_h^*)$ specified in Examples 5.70-5.75

Example	Grid X^h	Full GS	Active GS	CPU time	$\mathcal{Z}_h$
Ex. 5.70	51×51	3 381 300	13 402	28 sec.	2.1166
(Fig. 5.6)	101×101	52 025 100	61 022	3 min. 37 sec.	2.1166
	201×201	816 100 200	236 228	23 min. 35 sec.	2.1166
Ex. 5.71	51×51	3 381 300	12 784	28 sec.	2.7442
(Fig. 5.7)	101×101	52 025 100	30 386	3 min. 03 sec.	2.7442
	201×201	816 100 200	305 180	23 min. 44 sec.	2.7442
Ex. 5.72	51×51	3 381 300	19 618	30 sec.	1.1713
(Fig. 5.8)	101×101	52 025 100	104 148	3 min. 44 sec.	1.1714
	201×201	816 100 200	612 082	35 min. 46 sec.	1.1715
Ex. 5.73	51×51	3 381 300	13 665	24 sec.	0.30898
(Fig. 5.9)	101×101	52 025 100	71 609	2 min. 26 sec.	0.30863
	201×201	816 100 200	392 812	24 min. 44 sec.	0.30840
Ex. 5.74	53×53	3 943 836	13 274	33 sec.	1.7527
(Fig. 5.10)	101×101	52 025 100	56 288	3 min. 50 sec.	1.7524
	201×201	816 100 200	251 517	31 min. 02 sec.	1.7522
Ex. 5.75	61×61	$\approx 60\cdot10^6$	14 905	28 sec.	0.15795
(Fig. 5.11)	121×121	$\approx 125\cdot10^6$	71 313	4 min. 10 sec.	0.15770
	241×241	$\approx 600\cdot10^6$	352 624	29 min. 36 sec.	0.15757

Example 5.70. (Four point loads) For a square domain $\Omega = \big(-\frac{1}{2}, \frac{1}{2}\big)^2$ the load $f = \sum_{i=1}^4 \delta_{x_i}$ consists of four symmetrically spaced point forces as in Fig. 5.6(a): $x_i = (\pm\alpha, \pm\alpha)$ with $\alpha = 0.2$ (the forces are pointing upwards hence the symbol $\odot$). The pair of conic quadratic programs is solved for a regular grid X^h in accordance with (5.115) for three resolutions: $h \in \{1/50, 1/100, 1/200\}$. The finest resolution gives a 201×201 nodal grid X^h and thus $m = \#(X\backslash\partial\Omega) = 199^2 = 39\,601$ nodes in the interior Ω. The fully connected ground structure (abbrev. "Full GS" in Table 5.1) counts $N = \bar{m}\,(\bar{m}-1)/2 = 816\,100\,200$ potential members, where $\bar{m} = 201^2$. There is no need to discretize the load, i.e. $f_h = f$ and the load vector $\mathbf{f} \in \mathbb{R}^m$ has four non-zero components each equal to 1.

For $h = 1/200$ the adaptive member-adding algorithm solving the conic program $(\boldsymbol{P}_h)$, $(\boldsymbol{P}_h^*)$ converged in 10 iterations and the total computational time was around 24 minutes; the number of active members (abbrev. "Active GS" in Table 5.1) in the last iteration

was $N_{10} = 236\,228$. The objective value thus numerically obtained was $\mathcal{Z}_h = 2.1166$. The computational details, also for other resolutions h, are listed in Table 5.1.



(a) (b) (c)

Figure 5.6: Numerical prediction of optimal membrane subjected to four point forces: (a) optimal $\varsigma_{\mathscr{S}_h}$ (and an alternative solution in the top right corner); (b) optimal ϑ_{π_h}; (c) optimal u_h.

Fig. 5.6(c) shows the graph of interpolation u_h of nodal out-of-plane displacements $\mathbf{u}$ computed for resolution $h = 1/100$. Over the central square $Q := \mathrm{co}(\{x_i\})$ we can observe a plateau, i.e. numerically u_h is constant and approximately equals 0.529 in this part of Ω. Moreover the vector function w_h (not displayed in the figure) can be recognized as identity mapping id in Q therefore numerically $\frac{1}{2}\nabla u_h \otimes \nabla u_h + e(w_h) = e(w_h) = \mathrm{I}$ therein. This explains the pre-stress field $\varsigma_{\mathscr{S}_h}$ showed in Fig 5.6(a): it is composed of eight strings connecting the force application points x_i to points a_{ij} on the boundary and of huge number of strings in the square Q. It is so since conditions (iii), (iv) in (5.113) (or their counterparts in (5.99)) are trivially satisfied for any string k that is contained in the inner square. Ultimately only the equilibrium conditions (i) determine $\mathbf{s}$ therein and $\varsigma_{\mathscr{S}_h} \llcorner Q$ as a result. This gives plenty of possible optimal solutions and the interior point method implemented in MOSEK tends to give an intermediate solution that is a convex combination of all of those optimal configurations. The very same phenomenon was observed analytically in Example 5.46. A much simpler solution that is composed of 12 strings in total may be numerically provoked – its miniature is showed in the top right corner of Fig. 5.6(a). Another story may be told about the transverse force field ϑ_{π_h} that is displayed in Fig. 5.6(b): due to condition (iv) in (5.113) (or in (5.99)) non-zero transverse forces may occur only in strings that admit a slope of u_h. Consequently ϑ_{π_h} charges the eight strings going to the boundary only. It must be stressed that the arrows in Fig. 5.6(b) indicate the direction that is opposite to direction of $\vartheta_{\pi_h} \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^2)$ (which will be explained in Section 5.6).

Let us now give some details about the solutions (u_h, w_h) and $(\pi_h, \mathscr{S}_h)$ in the simpler case when the optimal system is composed of 12 strings, namely $\mathrm{spt}\,\mathscr{S}_h \subset G := \big(\bigcup_{ij}\{a_{ij}\} \cup \bigcup_i \{x_i\}\big)^2$ while positions of points a_{ij} are specified in Fig. 5.6(a). Let us stress that solution $(\pi_h, \mathscr{S}_h)$ is identical for the three resolutions $h \in \{1/50, 1/100, 1/200\}$ and it is expected to be exact. Below we give analytically computed values (justification will be given below)

that coincide with the numerical results. The values of (u_h, w_h) at the forces application points x_i are:

$$u_h(x_i) = \frac{\sqrt{7}}{5}, \qquad w_h(x_i) = x_i \qquad \forall\, i;$$

recall that (presumably) $w_h = \mathrm{id}$ and $u_h \equiv \frac{\sqrt{7}}{5} \approx 0.529$ on the whole Q. For the eight strings connecting points x_i to the boundary we have

$$\pi_h\Big(\big\{(x_i, a_{ij})\big\}\Big) = -\frac{1}{2}, \qquad \mathscr{S}_h\Big(\big\{(x_i, a_{ij})\big\}\Big) = \sqrt{\frac{5}{56}} \qquad \forall\, i, j;$$

note that equivalently we could redefine π_h so that $\pi_h\big(\big\{(a_{ij}, x_i)\big\}\big) = \frac{1}{2}$ (the order of x_i, a_{ij} must be swapped), see Remark 5.49. For the four strings that interconnect the force application points we have

$$\pi_h\Big(\big\{(x_i, x_{i+1})\big\}\Big) = 0, \qquad \mathscr{S}_h\Big(\big\{(x_i, x_{i+1})\big\}\Big) = \frac{1}{2\sqrt{7}} \qquad \forall\, i$$

where we agree that $x_5 := x_1$.

Considering that $(\pi_h, \mathscr{S}_h) \in \Xi(\Omega, f)$ the values $\pi_h\big(\big\{(x_i, a_{i1})\big\}\big) = -\frac{1}{2}$ are clear since necessarily $-\pi_h\big(\big\{(x_i, a_{i1})\big\}\big) - \pi_h\big(\big\{(x_i, a_{i2})\big\}\big) = f(x_i) = 1$ by property (i) in (5.89), i.e. by the out-of-plane equilibrium of the joints x_i. For given geometry of the string system one can check that the ratio between pre-stressing forces $\sqrt{\frac{5}{56}}$ and $\frac{1}{2\sqrt{7}}$ guarantees the in-plane equilibrium of the joints x_i, i.e. the property (ii) in (5.89). This way we have verified the first optimality condition (i) in (5.99). Next we show that optimality conditions (iii), (iv) in (5.99) hold true as well. For the eight strings going to the boundary we obtain:

$$\begin{aligned}
&\frac{1}{2}\,|u_h(a_{ij}) - u_h(x_i)|^2 + \langle w_h(a_{ij}) - w_h(x_i), a_{ij} - x_i\rangle \\
&= \frac{1}{2}\left|0 - \tfrac{\sqrt{7}}{5}\right|^2 + \langle 0 - x_i, a_{ij} - x_i\rangle = \frac{1}{10} = |a_{ij} - x_i|^2 \qquad \forall\, i, j,
\end{aligned}$$

while for the other four:

$$\begin{aligned}
&\frac{1}{2}\,|u_h(x_{i+1}) - u_h(x_i)|^2 + \langle w_h(x_{i+1}) - w_h(x_i), x_{i+1} - x_i\rangle \\
&= \frac{1}{2}\left|\tfrac{\sqrt{7}}{5} - \tfrac{\sqrt{7}}{5}\right|^2 + \langle x_{i+1} - x_i, x_{i+1} - x_i\rangle = |x_{i+1} - x_i|^2 \qquad \forall\, i
\end{aligned}$$

and condition (iii) follows. Then, if for the eight strings we define

$$\alpha_h(x_i, a_{ij}) := \frac{u_h(a_{ij}) - u_h(x_i)}{|a_{ij} - x_i|} = -\sqrt{\frac{14}{5}} \qquad \forall\, i, j$$

then it is straightforward to check that

$$\pi_h\Big(\big\{(x_i, a_{ij})\big\}\Big) = \alpha_h\big(x_i, a_{ij}\big)\, \mathscr{S}_h\Big(\big\{(x_i, a_{ij})\big\}\Big) \qquad \forall\, i, j \qquad (5.116)$$

hence the optimality condition (iv). To prove that the pairs (u_h, w_h) and $(\pi_h, \mathscr{S}_h)$ are exact solutions it is left to show the condition (ii) in (5.99), i.e. the two-point inequality condition that holds for any pair $(x, y) \in \overline{\Omega} \times \overline{\Omega}$. This part is difficult as the numerical method delivers (u_h, w_h) on X^h only, while on $\overline{\Omega}$ we use piece-wise affine interpolation which may be inaccurate and it is also difficult to work with analytically. The best way seems to consist in proposing a Lipschitz continuous extension $(\bar{u}, \bar{w})$ of functions (u, w)

defined on $Q \cup \partial\Omega$, i.e. $u \equiv \frac{\sqrt{7}}{5}$, $w = \mathrm{id}$ on Q and $(u, w) = (0, 0)$ on $\partial\Omega$. Then instead of checking the two-point condition (ii) in (5.99) one can validate the differential condition (ii) in (5.67), i.e. $\frac{1}{2}\,\nabla\bar{u} \otimes \nabla\bar{u} + e(\bar{w}) \preceq \mathrm{I}$ in $\overline{\Omega}$. Until this moment the author has failed to find such an extension.

Although we are not able to show the exact optimality of the numerical solution we can prove that if the 12-string topology is assumed the positions of the boundary points a_{ij} in Fig. 5.6(a) are in fact optimal. We will extend our argument to other forces' positions $x_i = (\pm\alpha, \pm\alpha)$. To that aim consider the two-point condition for pairs (x_1, a) where $a = (\frac{1}{2}, \beta) \in \left[(\frac{1}{2}, -\frac{1}{2}), (\frac{1}{2}, \frac{1}{2})\right] \subset \partial\Omega$ is an arbitrary point lying on the right edge of the square Ω. Again we fix $w_h(x_1) = x_1 = (\alpha, \alpha)$ while freeing the value $u_h(x_1)$ to find that:

$$\max_{a \in \left[(\frac{1}{2}, -\frac{1}{2}), (\frac{1}{2}, \frac{1}{2})\right]} \left\{ |a - x_1|^2 - \tfrac{1}{2}\,|u_h(a) - u_h(x_1)|^2 - \langle w_h(a) - w_h(x_1), a - x_1 \rangle \right\}$$
$$= \max_{\beta \in \left[-\frac{1}{2}, \frac{1}{2}\right]} \left\{ \beta^2 - \alpha\,\beta + \tfrac{1}{4}(1 - 2\alpha) - \tfrac{1}{2}\,|u_h(x_1)|^2 \right\} = \tfrac{1}{4}(1 - 2\alpha - \alpha^2) - \tfrac{1}{2}\,|u_h(x_1)|^2.$$

The latter maximum was found for $\hat{\beta} = \frac{\alpha}{2}$. We conclude that the two-point condition is not violated for any pair (x_1, a) with a lying on the right edge if it is not violated for the pair $(x_1, \hat{a})$ where

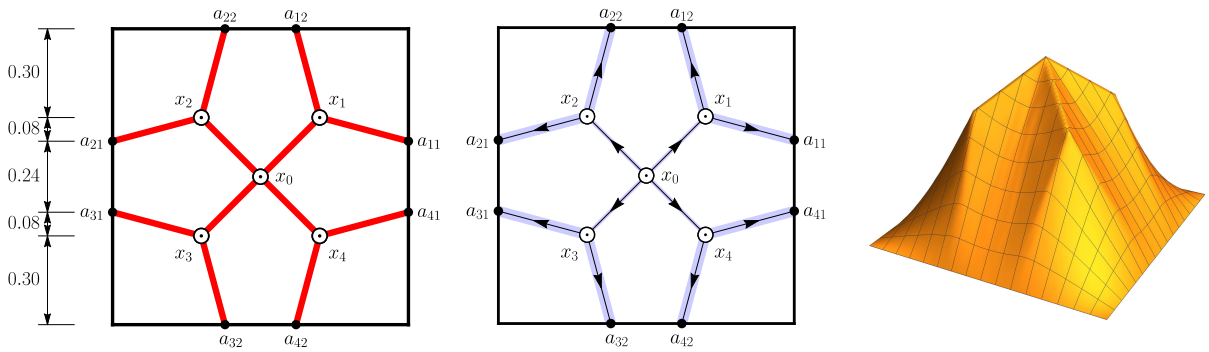
$$\hat{a} = (\tfrac{1}{2}, \hat{\beta}) = (\tfrac{1}{2}, \tfrac{1}{2}\alpha).$$

Furthermore, the condition will be sharp for this pair if $|u_h(x_1)| = \sqrt{1/2 - \alpha - \alpha^2/2}$. Once we take $\alpha = \frac{2}{10}$ as in the numerical solution we obtain $\hat{a} = (\frac{1}{2}, \frac{1}{10})$ which is exactly a_{11} while $|u_h(x_1)| = \frac{\sqrt{7}}{5}$. Analogous reasoning may be repeated for other x_i and other square's edges. We infer that the numerically obtained positions of boundary points a_{ij} are indeed optimal therefore they will not change with subsequent refinements of the grid X^h. In the process we have also explained how the "exact" value $u_h(x_i) = \frac{\sqrt{7}}{5}$ was calculated. Based on this value, through condition (5.116), one deduce the values of $\mathscr{S}$ on the eight strings. We have also predicted how the solution will change with the position α of the load, which was confirmed numerically. Note, however, that the 12-string configuration is no longer optimal if $\alpha \geq \frac{1}{3}$. In that case the optimal system disintegrates into four independent strings, each oriented as in Fig. 5.5(c). △

Example 5.71 (**Five point forces**)**.** We modify the previous example by adding one point force in the centre of the square, i.e. $f = \delta_{x_0} + \sum_{i=1}^{4} \delta_{x_i}$ where $x_0 = (0, 0)$. Once again $f_h = f$ and Fig. 5.7 presents numerical solutions $\varsigma_{\mathscr{S}_h}$, ϑ_{π_h} and u_h. We can see that the plateau is no longer present in the graph of u_h and the 12-string structure $\varsigma_{\mathscr{S}_h}$ seems stable (with respect to perturbations of the algorithm). In contrast to Example 5.70 here the points a_{ij} where strings connect to the boundary are slightly shifted towards the square's corners. Computational details for three resolutions h may be found in Table 5.1.

We observe a remarkable feature of the optimal configuration that also characterized the 12-string solution in the previous example:

$$\mathrm{spt}\,(\pi_h, \mathscr{S}_h) \subset \big(\mathrm{spt}\, f \cup \partial\Omega\big)^2, \tag{5.117}$$



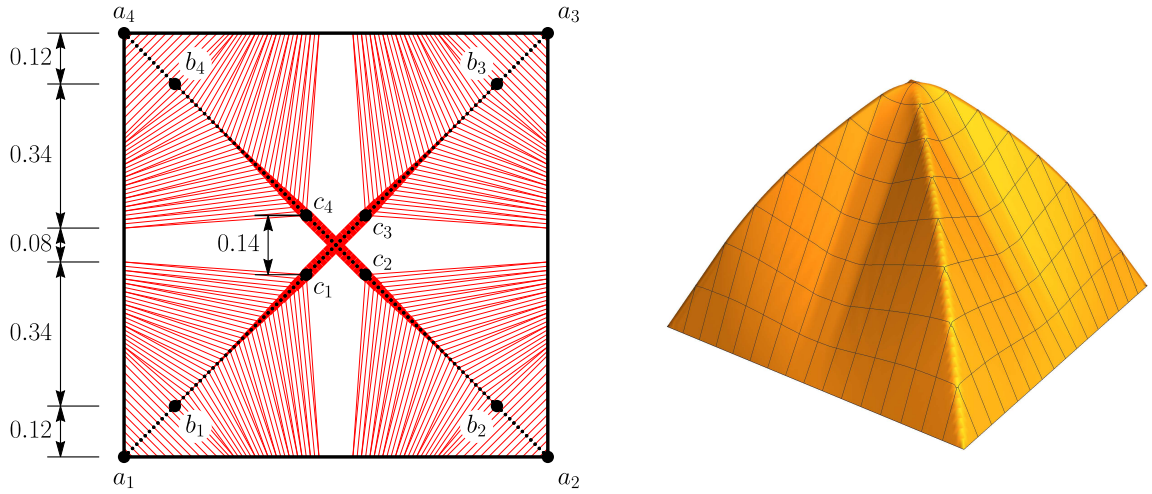
(a) (b) (c)

Figure 5.7: Numerical prediction of optimal membrane subjected to five point forces: (a) optimal $\varsigma_{\mathscr{S}_h}$; (b) optimal ϑ_{π_h}; (c) optimal u_h.

more precisely there exists at least one optimal string system whose joints are either boundary points or points of the forces application. This introduces a huge contrast with respect to truss approximations of Michell structures for which it is typical that the number of the extra engaged joints tends to infinity when $h \to 0$, see Fig. 4.9(a,b). Similar conclusions may be drawn for the exact solutions $(\pi, \mathscr{S})$ found for the one-point-force load f in Section 5.5.3. In fact, no exceptions from that rule were found in the numerical experiments, including the ones presented below. In Section 5.7.1 we shall convey a mathematical premise for the property (5.117) that will be a conjecture for existence of an exact solution $(\pi, \mathscr{S})$ of $(\widetilde{\mathscr{P}^*})$ when f is finitely supported. △

Example 5.72 (Diagonal load). We continue with a square domain $\Omega = \left(-\frac{1}{2}, \frac{1}{2}\right)^2$ while varying the load: this time around we consider a load that is concentrated along the square's diagonals, namely $f = \mathcal{H}^1 \llcorner [a_1, a_3] + \mathcal{H}^1 \llcorner [a_2, a_4]$, cf. Fig. 5.8(a) where also the discretized load f_h is showed in the form of tiny black dots aligned on the diagonals. Numerical computations were performed for three resolutions h (see Table 5.1 for details) while Fig. 5.8(a) shows solution $\varsigma_{\mathscr{S}_h}$ for the intermediate resolution $h = 1/100$; the graph of the interpolated function u_h is given in Fig. 5.8(b).

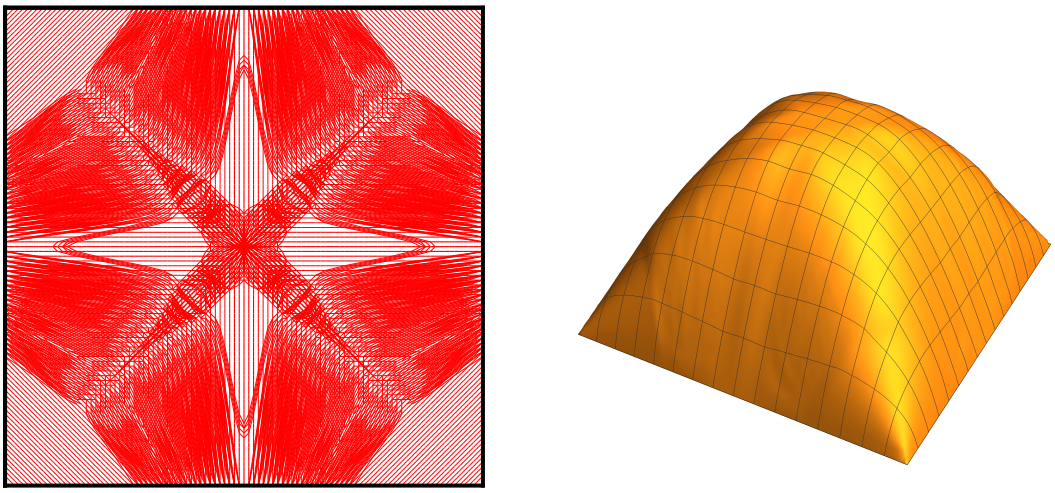
Similarly as in previous examples, the support of $(\pi_h, \mathscr{S}_h)$ is contained in $(\operatorname{spt} f_h \cup \partial\Omega)^2$, i.e. no intermediate points are essential for the optimal transmission of the load. The numerical results strongly indicate that exact ς that is optimal in $(\mathscr{P}^*)$ splits into an absolutely continuous part with respect to $\mathcal{H}^1 \llcorner [b_1, b_3] + \mathcal{H}^1 \llcorner [b_2, b_4]$ and into an absolutely continuous part with respect to Lebesgue measure restricted to four quadrilaterals. The two 1D diagonal strings have cross-sectional area that varies e.g. from zero area at b_1 to the maximal area at c_1, then it is constant from c_1 to c_3 and it varies again from c_3 to b_3. The area variation is due to the "continuous" fans of strings that are inclined with respect to the diagonals. In both parts the optimal ς (or rather $\sigma = \frac{d\varsigma}{d\mu}$ with $\mu = \operatorname{Tr}\varsigma$) is expected to be rank one. Based on Fig. 5.8(b) we may predict that the restrictions of $u|_{[b_1,b_3]}$ and $u|_{[b_2,b_4]}$ are strictly concave hence we expect that the problem $(\widetilde{\mathscr{P}^*})$ with the initial load f has no solution (see Remark 5.59). △



(a) (b)

Figure 5.8: Numerical prediction of optimal membrane subjected to the load f distributed along the square's diagonals (being discretized to f_h that is illustrated as small black dots): (a) optimal $\varsigma_{\mathscr{S}_h}$; (b) optimal u_h.

Example 5.73 (**Pressure load**)**.** For the square domain $\Omega = \left(-\frac{1}{2}, \frac{1}{2}\right)^2$ we now consider a uniformly distributed pressure load, namely $f = \mathcal{L}^2 \llcorner \Omega$. The discretized load f_h is effectively charging every single node in X^h. Solutions $\varsigma_{\mathscr{S}_h}$ (for an intermediate resolution $h = 1/100$, see Table 5.1 for more details) and u_h are showed in Fig. 5.9(a) and (b), respectively.



(a) (b)

Figure 5.9: Numerical prediction of optimal membrane subjected to a uniform pressure load $f = \mathcal{L}^2 \llcorner \Omega$: (a) optimal $\varsigma_{\mathscr{S}_h}$; (b) optimal u_h.

The prediction $\varsigma_{\mathscr{S}_h}$, that approximates an exact solution $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$, is difficult to analyse. Surely $\mathrm{spt}(\varsigma) = \overline{\Omega}$ and it seems that $\varsigma \ll \mathcal{L}^2$: there is a concentration of big number of thick strings in the very centre of the square domain, although the author's

guess is that they approximate a function $\sigma = \frac{d\varsigma}{d\mathcal{L}^2} \in L^1(\Omega; \mathcal{S}^{2\times 2})$ that blows up to infinity there (cf. Example 5.45). The exact pre-stress field ς is also suspected to be rank-one in a large portion of the domain except in the neighbourhood of diagonals, far from the corners, where ς seems to be of full rank (the presumed rank-two region stands out visually in Fig. 5.9(a)).

In the "rank-two region" the orientation of strings seem to be quite arbitrary – we are not observing "quasi orthogonal" nets of strings known from truss approximations of Michell structures, see Fig 4.9 again. On top of that the string configuration in the rank-region is very unstable with respect to perturbation of the numerical algorithm, i.e. perturbations that for instance favoured the 12-string solution in Example 5.70. At the same time we do not see a blur of all possible directions of strings once the algorithm is not perturbed at all, cf. Fig. 5.6(a). Behind this peculiar behaviour of the rank-two region there is an explanation that partially goes through analysis of the stress tensor σ at a point and partially through analysis of an approximative string system.

First of all, the orthogonal nets of material fibres – the so called *Hencky nets* – occur in Michell structures in the tension-compression regions, i.e. when (in 2D case) $\det\sigma < 0$. This finds its reflection in the analysis of the Fibrous Material Design setting of the (FMD) problem that in Section 4.7.2 was proved to be equivalent to Michell problem. In Example 3.18, in the Case a) (when $\det\sigma < 0$) we have shown that the set of optimal "fibrous" Hooke tensors $\underline{\mathcal{K}}_1(\sigma) \subset \mathcal{K}_{fib}$ is a singleton whose unique element is $\underline{H}_\sigma = \sum_{i=1}^2 \alpha_i\, v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma)\otimes v_i(\sigma)$ where $v_i(\sigma)$ are the mutually orthogonal eigenvectors of σ and α_i are proportional to absolute values of the corresponding eigenvalues of σ. Of course this scenario can never occur in the optimal membrane design since always $\sigma \succeq 0$. With this restriction we move on to the Canvas-like Material Design setting of (FMD) problem. In Example 3.28 we have learned that once $\mathrm{rank}(\sigma) = d = 2$ the set of optimal Hooke tensors reads

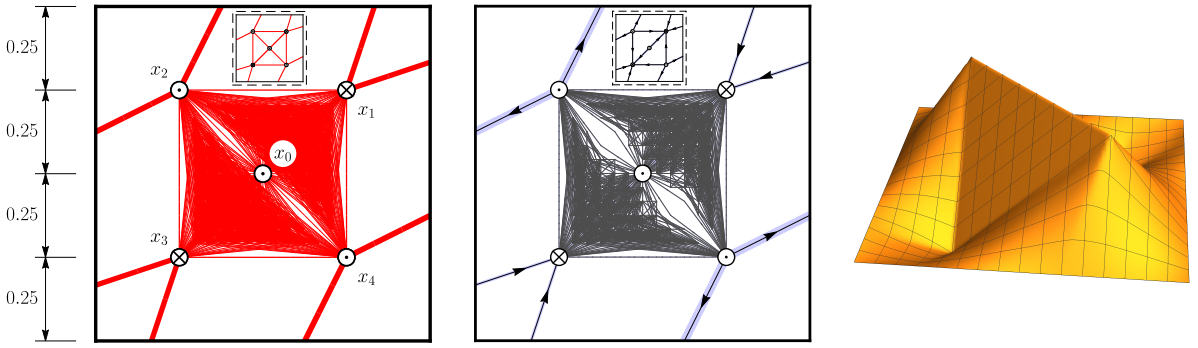
$$\underline{\mathcal{K}}_{1,+}(\sigma) = \left\{ \sum_{i=1}^m \alpha_i\, \tau_i\otimes\tau_i\otimes\tau_i\otimes\tau_i \;:\; \tau_i \in \operatorname{Im}\sigma \cap S^{d-1},\ \sum_{i=1}^m \alpha_i = 1,\ \frac{\sigma}{\operatorname{Tr}\sigma} = \sum_{i=1}^m \alpha_i\, \tau_i\otimes\tau_i \right\}.$$

In Remark 3.29 it was pointed out that the set $\underline{\mathcal{K}}_{1,+}(\sigma)$ is very rich once $\mathrm{rank}(\sigma) > 1$ and that mutual orthogonality of τ_i is not obligatory, in particular τ_i do not have to be parallel to eigenvectors of σ (even if σ is not isotropic). Globally we design a Hooke field $\lambda = \mathscr{H}\,\mathcal{L}^2 \llcorner \Omega$ such that $\frac{\mathscr{H}(x)}{\operatorname{Tr}\mathscr{H}(x)} \in \underline{\mathcal{K}}_{1,+}\big(\sigma(x)\big)$ a.e. If we are to trust Conjecture 5.69, i.e. that the 1D field $\lambda_{\mathscr{S}_h}$ supported on the ground structure (see (5.114)) weakly-* converges to such λ, then it is understandable that in the rank-two region the strings are oriented pretty much arbitrarily.

It is still not yet clear why the rank-two regions are not numerically represented by blurry clouds of strings will all possible orientations as in Fig. 5.6(a). In turn this cannot be interpreted by analysing optimality of Hooke tensor H for the exact solution σ at a point – there are simply no other conditions for if than $H \in \underline{\mathcal{K}}_{1,+}(\sigma)$. This time we must work with the numerical approximation by a string system and the optimality conditions (5.113) that it must satisfy. In particular there is condition (iv), i.e. $\mathbf{p}(k) = \frac{(\mathbf{Hu})(k)}{\mathbf{L}(k)}\,\mathbf{s}(k)$

must hold for each string, which combined with the out-of-plane equilibrium condition in (i): $\mathbf{H}^\top \mathbf{p} = \mathbf{f}$ introduces additional constraints on the configuration of the strings in the rank-two regions. In this example the behaviour of the interpolated function u_h is complicated in the rank-two regions, i.e. it appears to be strictly concave (it is not affine as in the next example). This is connected to the fact that there is a vertical force applied at every single node in X^h, i.e. $\mathbf{f}$ has no zero entries. In Example 5.70 the rank-two region was the central square where u_h was constant and the two aforementioned conditions are met trivially there. In this square the only binding optimality conditions for vector $\mathbf{s}$ were the in-plane equilibrium equations $\mathbf{B}_1^\top \mathbf{s} = \mathbf{0}$, $\mathbf{B}_2^\top \mathbf{s} = \mathbf{0}$. This made much more optimal string configurations possible, hence the blur in Fig. 5.6(a). △

Example 5.74 (**Signed load**)**.** For the same square domain we investigate the case of a signed discrete load: $f = \delta_{x_0} - \delta_{x_1} + \delta_{x_2} - \delta_{x_3} + \delta_{x_4}$, see Fig. 5.10 where negative point forces – i.e. forces pointing downwards – are marked by $\otimes$ symbol and numerical solutions $\varsigma_{\mathscr{S}_h}$, ϑ_{π_h}, u_h are showed. Computational details for three resolutions h are to be found in Table 5.1.



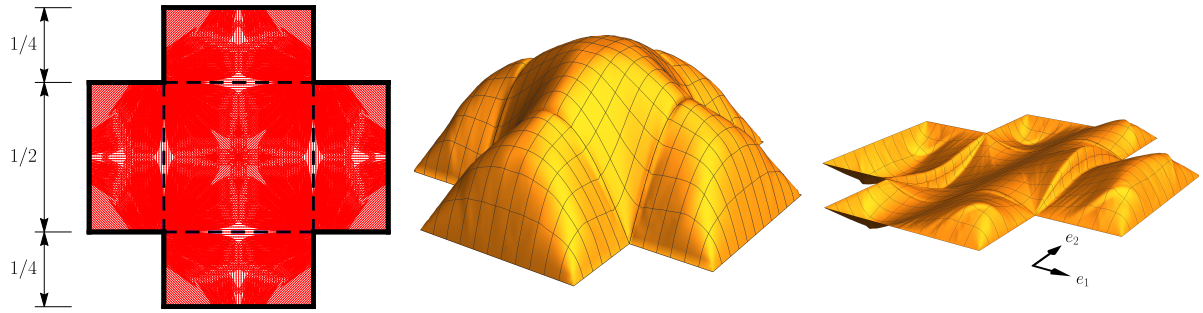
(a) (b) (c)

Figure 5.10: Numerical prediction of optimal membrane subjected to five-point signed load: (a) optimal $\varsigma_{\mathscr{S}_h}$ (and an alternative solution); (b) optimal ϑ_{π_h} (and an alternative solution); (c) optimal u_h.

We can see that u_h is affine within two triangles $\mathrm{co}\big(\{x_2, x_3, x_4\}\big)$ and $\mathrm{co}\big(\{x_1, x_2, x_4\}\big)$ while w_h (not displayed here) turns out to be affine in the whole central square $\mathrm{co}\big(\{x_1, x_2, x_3, x_4\}\big)$ where it (according to numerical prediction) satisfies $\frac{1}{2}\nabla u_h \otimes \nabla u_h + e(w_h) = \mathrm{I}$. Accordingly, no point-wise constraint is prescribed for $\varsigma_{\mathscr{S}_h}$ in this square and multiple solutions may appear as in the Example 5.70. However, contrarily to the latter example $\nabla u_h \neq 0$ in the square, therefore we have a high non-uniqueness of ϑ_{π_h} as well. In turn, since u_h is affine in the two triangles the optimality conditions (i), (iv) are simpler to meet numerically than it was in the case of the rank-two regions in Example 5.73 therefore there are more optimal string configurations, which explains the blurred regions in Fig. 5.10(a,b). In the top of these figures equivalent solutions $\varsigma_{\mathscr{S}_h}$, ϑ_{π_h} are shown, which are provoked by numerical perturbation of the algorithm. In the end the numerical simulation once again suggests that there exists an optimal string system that does not employ nodes that are not loaded and do not lie on the boundary, i.e. $\mathrm{spt}\,(\pi_h, \mathscr{S}_h) \subset \big(\mathrm{spt}\, f \cup \partial\Omega\big)^2$. △

In all the hitherto discussed examples the numerical solution w_h, although not displayed, appears to approximate a function w that is continuous, meanwhile from characterization of the set $\overline{\mathcal{X}}_1$ in Theorem 5.32 we know that we may expect that the exact solution is merely in the space $BV \cap L^\infty(\Omega;\mathbb{R}^2)$. We can make an even stronger suspicion: all solutions (u_h, w_h) so far seem to approximate a Lipschitz continuous pair (u, w). The only counter-example to this tendency was found by the author through visiting a non-convex domain Ω. Recall that in Remark 5.52 we have outlined how to adapt the theory of optimal string system to the case of non-convex design domains. While using the numerical method by ground structure, however, the construction in Remark 5.52 cannot be employed directly, simply because not every point on $\partial\Omega$ belongs to X^h. As a result, in practise it could happen that strings of non-zero cross section go out of a non-convex polygon $\overline{\Omega}$ in close vicinity of its re-entrant corners. The idea to remedy this is simple enough: the strings (x, y) with $[x, y] \not\subseteq \overline{\Omega}$ may be *a priori* removed from the ground structure. This is very reasonable from the computational point of view as we eliminate a huge number of constraints or variables in problems $(\boldsymbol{P}_h)$ or $(\boldsymbol{P}_h^*)$, respectively. Such string elimination has been already implemented by the present author for optimization of trusses, see [31].

Example 5.75 (A non-convex cross-shaped domain). We choose a non-convex domain Ω in a shape of a symmetric cross; the load is again uniformly distributed, namely $f = \mathcal{L}^2 \llcorner \Omega$. Solutions $\varsigma_{\mathscr{S}_h}$ (for an intermediate resolution $h = 1/100$) and u_h are showed in Fig. 5.9(a) and (b), respectively. Fig. 5.11(c) shows the graph of $\langle w_h, \mathrm{e}_1 \rangle$, i.e. of one of the components of $w_h : \overline{\Omega} \to \mathbb{R}^2$. The computational data, also for other resolutions of h, is listed in Table 5.1 (since Ω is non-convex the precise number of members in the full ground structure would require heavy computing, instead we give a rough estimate).



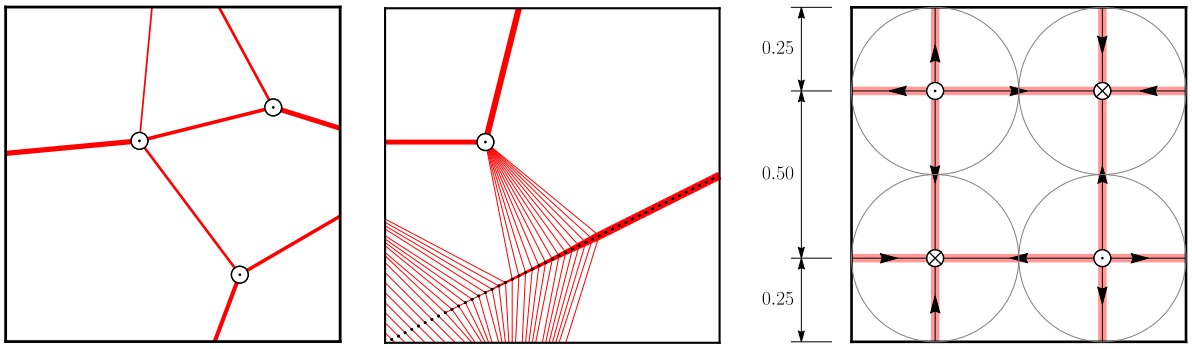
(a) (b) (c)

Figure 5.11: Numerical prediction of optimal membrane in a cross-shaped domain Ω and subjected to a uniformly distributed pressure load $f = \mathcal{L}^2 \llcorner \Omega$: (a) optimal $\varsigma_{\mathscr{S}_h}$; (b) optimal u_h; (c) the component $\langle w_h, \mathrm{e}_1 \rangle$ of optimal w_h.

We observe that the string system in the central square connecting the four re-entrant corners is disconnected from the rest of the remaining four parts, namely there are no strings that interconnect the five regions, cf. the dashed lines partitioning the domain in Fig. 5.11(a). The numerical prediction u_h in Fig. 5.11(b) suggests that the exact solution u is Lipschitz continuous. The same cannot be inferred about the exact optimal function

w based on Fig. 5.11(c) which clearly indicates that w is discontinuous along the four aforementioned lines that divide Ω. More precisely it is only the component w_ν, being normal with respect to those lines, that admits a jump – note that $\langle w, \mathrm{e}_1 \rangle$ appears to be continuous along the two of those lines that are parallel to e_1. In the end the numerical simulation confirms that $w \in BV \cap L^\infty(\Omega; \mathbb{R}^2)$ and $\epsilon_s(w) \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{2\times 2}_-)$ i.e. the singular part of the distributional derivative $\epsilon(w)$ is a negative semi-definite measure, cf. condition (iii) in Theorem 5.32. △

Example 5.76 (**Miscellaneous**)**.** We conclude the presentation with another three solutions for the square domain Ω: Fig. 5.12(a,b) show solutions $\varsigma_{\mathscr{S}_h}$ for asymmetric positive loads: three point forces $f = \sum_{i=1}^3 \delta_{x_i}$ in Fig. 5.12(a) and sum of a point force and load distributed along a line $f = \delta_{x_0} + \mathcal{H}^1 \llcorner [a, b]$ in Fig. 5.12(b). Position of the signed load



(a) (b) (c)

Figure 5.12: Numerical predictions: (a) optimal $\varsigma_{\mathscr{S}_h}$ for three asymmetric point forces; (b) optimal $\varsigma_{\mathscr{S}_h}$ for point force and force distributed along a line; (c) optimal $\varsigma_{\mathscr{S}_h}$ and ϑ_{π_h} for a signed load.

$f = \sum_{i=1}^2 \delta_{x_i} - \sum_{i=3}^4 \delta_{x_i}$ in Fig. 5.12(c) was strategically picked: the solution w_h turns out to be zero, which will have some interesting implications in the context of Section 5.6.1, cf. Remark 5.79. The figure represents both optimal $\varsigma_{\mathscr{S}_h}$ and ϑ_{π_h} (note the arrows). The solutions may be proved to be exact by a simple adaptation of the proof of Proposition 5.60: the graph of function u can be constructed as an extension by zero of the four cones of revolution with disks marked in Fig. 5.12(c) as bases and vertices $\big(x_i, \operatorname{sign}\big(f(x_i)\big)\sqrt{2}/4\big)$. The key point is that in-flows and out-flows of ϑ_{π_h} are matched at points where the pairs of disks touch and thus they play the role of additional points in the Dirichlet region.

For all three examples herein solved numerically we can once more deduce that $\operatorname{spt}(\pi_h, \mathscr{S}_h) \subset \big(\operatorname{spt} f_h \cup \partial\Omega\big)^2$. △

5.6. Maximal Monge-Kantorovich metrics versus optimal design of membranes

In the very first section of this chapter we have introduced the optimal membrane problem (OPM) (in its plastic setting) by modifying the Free Material Design problem

(FMD) posed for the stationary heat conduction, namely:

$$\mathcal{C}_{\min} = \min\left\{\mathcal{C}(\lambda) \ : \ \lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+), \ \int \operatorname{Tr}\lambda \leq C_0\right\} \tag{FMD}$$

where $\mathcal{C}(\lambda)$ depends on the source/load f, see (5.1). The alteration consisted in an extra constraint for the design variable being a matrix valued measure:

$$\mathcal{C}_{\min,0} = \min\left\{\mathcal{C}(\varsigma) \ : \ \varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+), \ \int \operatorname{Tr}\varsigma \leq C_0, \ \ \operatorname{Div}\varsigma = 0 \ \text{ in } \Omega\right\} \tag{OPM}$$

The zero divergence condition $\operatorname{Div}\varsigma = 0$ in $\mathcal{D}'(\Omega;\mathbb{R}^d)$ changed the physical interpretation of the problem entirely: the conductivity field λ became a pre-stress field ς that constituted a membrane. As it unravelled throughout this chapter the mathematical approach to the problem had to be heavily adjusted as well.

Throughout the entire Section 5.6 we shall assume that the design domain Ω is convex. In Section 4.8 we have demonstrated how the (FMD) problem for the stationary heat conduction, i.e. in the scalar variant where $u : \Omega \to \mathbb{R}$, can be tackled by means of the methods of optimal transport. Assuming the Dirichlet region to be $\Upsilon_0 = \partial\Omega$, for a transportation plan $\gamma \in \mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$ solving the associated problem ($\mathrm{OTP}^{\partial\Omega}$) the optimal conductivity field $\lambda = \lambda_\gamma$ was to be constructed via Theorem 4.67. The first question that we address in this section is as follows: what criteria the load f must meet so that $\varsigma_\gamma := \lambda_\gamma$ solves the optimal membrane problem (OPM) as well, namely when $\operatorname{Div}\lambda_\gamma = 0$ in Ω? A somewhat parallel question reads: for which f the objective values $\mathcal{C}_{\min}$ and $\mathcal{C}_{\min,0}$ coincide? In Section 5.6.1 below we shall give sufficient conditions for a non-negative load $f \in \mathcal{M}_+(\overline{\Omega})$ to meet both requirements.

The foregoing conditions for equivalence between problems (FMD) and (OPM) will turn out to be restrictive – for general f the link between (OPM) and the optimal transport problem as we know it from Section 4.8 breaks down. Nonetheless, since in problem ($\overline{\mathscr{P}}$) functions (u, w) are constrained by a two-point condition, we will manage to find a way to Kantorovich-Rubinstein duality. By exploiting the characterization of the set $\overline{\mathcal{X}}_1$ in Theorem 5.32 together with Remark 5.33 equality $\mathcal{Z} = \sup\overline{\mathscr{P}}$ can be rewritten as:

$$\mathcal{Z} = \sup_{\mathbf{v}\in\mathbf{V}_\Omega}\left\{\sup_{u\in C_0(\Omega)}\left\{\int u\,df \ : \ u(y) - u(x) \leq \ell_{\mathbf{v}}(x,y) \quad \forall\,(x,y)\in\overline{\Omega}\times\overline{\Omega}\right\}\right\} \tag{5.118}$$

where for each $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ we define

$$\ell_{\mathbf{v}}(x,y) := \min\left\{\sqrt{2\langle y'-x', y-x\rangle} \ : \ x' \in \mathbf{v}(x), \ \ y' \in \mathbf{v}(y)\right\}. \tag{5.119}$$

First of all we note that each multifunction in $\mathbf{V}_\Omega$ is a maximal monotone map hence the term under the square root is always non-negative. One can see that the two-point condition $u(y)-u(x) \leq \ell_{\mathbf{v}}(x,y)$ is equivalent to (5.53). The inner supremum in (5.118) is reminiscent of the one in Lemma 4.66, the only difference is that the Euclidean distance $|y-x|$ is now replaced by $\ell_{\mathbf{v}}(x,y)$. In order to make use of Kantorovich-Rubinstein theorem (cf. 4.59), that is to write the inner supremum as a minimum involving transport optimization, we require that $\ell_{\mathbf{v}}$ satisfies the triangle inequality, see Remark 4.60. For general $\mathbf{v} \in \mathbf{V}_\Omega$ this fails to be true. With the help of the maximal-monotone-map framework in Section 5.6.2 we

will construct a cost function $c_{\mathbf{v}} : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}_+$ that for a given $\mathbf{v}$ is the largest symmetric and sub-additive minorant of $\ell_{\mathbf{v}}$. It will turn out that the two-point condition in (5.118) will be equivalent to $u(y) - u(x) \leq c_{\mathbf{v}}(x, y)$, which will pave a way to Kantorovich-Rubinstein duality after all. With the inner supremum replaced by the Wasserstein distance a new role may be assigned to the variable $\mathbf{v}$: it induces a pseudo-metric $c_{\mathbf{v}}$ and it is to be chosen amongst $\mathbf{V}_\Omega$ so that the Monge-Kantorovich cost $W_{c_{\mathbf{v}}}^{\partial\Omega}(f_+, f_-)$ involving integral $\int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}(x, y)\, \gamma(dxdy)$ is maximal. From (5.118) we see that this maximal cost is equal exactly to $\mathcal{Z}$. The problem of finding "the worst $c_{\mathbf{v}}$ metric" will be studied in Section 5.6.3.

5.6.1. Sufficient conditions for equivalence between (FMD) and (OPM)

Confining ourselves to non-negative loads $f \in \mathcal{M}_+(\overline{\Omega})$ we will now give a sufficient condition under which the inequality

$$\mathcal{C}_{\min} = \min(\text{FMD}) \quad \leq \quad \mathcal{C}_{\min,0} = \min(\text{OPM}) \tag{5.120}$$

is an actual equality. It is enough to give criteria under which there exists a solution λ to (FMD) such that $\operatorname{Div}\lambda = 0$ in $\mathcal{D}'(\Omega; \mathbb{R}^d)$. To that aim we will employ the link between (FMD) and ($\text{OTP}^{\partial\Omega}$) that was established in Section 4.8.2 (note that we choose $\Upsilon_0 = \partial\Omega$ in the Free Material Design problem). Accordingly, for the six point source/load f in Example 4.71 the optimal measure $\lambda = \lambda_\gamma$, constructed via (4.73) based on the optimal transport plan γ as in Fig. 4.11(b), yields a non-zero divergence $\operatorname{Div}\lambda_\gamma$. On the other hand, for a one-point f applied at the very centre of the square Ω the optimal plan γ may be chosen as in Fig. 4.11 where the transport of mass f is uniformly partitioned into four rays going to different centres of the square's edges – the associated measure λ_γ is then of zero divergence.

For the square domain its centre turns out to be the only point where a positive load f can be applied in order that (5.120) is an equality. Below we shall prove that in case of arbitrary convex domain Ω it is enough that the support of f is contained in the set

$$M(\Omega) := \Big\{ x \in \Omega \ : \ x \in \operatorname{co}\big(\Pi_{\partial\Omega}(x)\big) \Big\}$$

where $\Pi_{\partial\Omega}(x)$ is the set of boundary point that are of maximal Euclidean distance with respect to x, cf. the definition in (4.74). The set $M(\Omega)$ is often referred to as *the high ridge of* Ω.

Proposition 5.77. *For a convex domain Ω assume that $f \in \mathcal{M}_+(\overline{\Omega})$ (f is non-negative) and $\operatorname{spt}(f) \subset M(\Omega)$. Then the* (FMD) *problem above admits a divergence free solution λ. As a result*

$$\mathcal{C}_{\min,0} = \mathcal{C}_{\min}.$$

Proof. Since the set $\mathcal{G}^{\partial\Omega} = \big\{(x, y) \in \overline{\Omega} \times \partial\Omega \ : \ y \in \Pi^{\partial\Omega}(x)\big\}$ is closed (see Section 4.8.2) one can easily show that the set $\mathscr{G} := \big\{(x, p) \in \operatorname{spt} f \times \mathrm{P}(\partial\Omega) : \operatorname{spt} p \subset \Pi_{\partial\Omega}(x)\big\}$ is closed as well (with $\mathrm{P}(\partial\Omega)$ endowed with the weak-* topology). With $[p] := \int_{\partial\Omega} y\, p(dy)$ standing for the barycentre of p we consider the subset:

$$\Big\{(x, p) \in \mathscr{G} \ : \ [p] = x\Big\} \tag{5.121}$$

which is again closed since the mapping $p \mapsto [p]$ is continuous. The set above is a graph of the multifunction $\Gamma : \operatorname{spt} f \to 2^{\mathrm{P}(\partial\Omega)}$. Non-emptiness of $\Gamma(x)$ for each $x \in \operatorname{spt} f$ follows by virtue of Choquet's theorem: indeed, since $\operatorname{spt} f \subset M(\Omega)$ every such x belongs to $\operatorname{co}\big(\Pi_{\partial\Omega}(x)\big)$. Then, Γ is upper semi-continuous and therefore it is a Borel measurable multifunction owing to Theorem 2.37 hence, by Theorem 2.36, it admits a Borel measurable selection $x \mapsto p^x$ (see Definition 2.26). We may thus define a transportation plan as a generalized product (see [8, Definition 2.27]) $\gamma = f \otimes p^x \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ for which the generalized Fubini theorem holds:

$$\langle \varphi, \gamma \rangle_{(C,\mathcal{M})} := \int_{\overline{\Omega}} \left(\int_{\overline{\Omega}} \varphi(x,y)\, p^x(dy) \right) f(dx) \qquad \forall\, \varphi \in C(\overline{\Omega}).$$

It is straightforward to show that $\operatorname{spt}\gamma \subset \mathcal{G}^{\partial\Omega}$ and $\pi_1 \# \gamma = f$ therefore γ solves the problem ($\mathrm{OTP}^{\partial\Omega}$) owing to Proposition 4.68. Consequently $\lambda_\gamma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ constructed in Theorem 4.64 solves the (FMD) problem by virtue of Theorem 4.67. To complete the proof we must show that $\operatorname{Div} \lambda_\gamma = 0$ in $\mathcal{D}'(\Omega; \mathbb{R}^d)$. For any $\phi \in \mathcal{D}(\Omega; \mathbb{R}^d)$ we have

$$\begin{aligned}
&\frac{Z}{C_0} \langle \operatorname{Div} \lambda_\gamma, \phi \rangle = - \int_{\overline{\Omega}\times\overline{\Omega}} \left(\int_{[x,y]} \langle \nabla\phi, \tau^{x,y} \otimes \tau^{x,y} \rangle\, \mathcal{H}^1 \right) \gamma(dxdy) \\
&= - \int_{\overline{\Omega}\times\overline{\Omega}} \left\langle \phi(y) - \phi(x), \frac{y-x}{|y-x|} \right\rangle \gamma(dxdy) = \int_{\overline{\Omega}} \left(\int_{\partial\Omega} \left\langle \phi(x), \frac{y-x}{|y-x|} \right\rangle p^x(dy) \right) f(dx) \\
&= \int_{\overline{\Omega}} \frac{\langle \phi(x), [p^x] - x \rangle}{d^{\partial\Omega}(x)}\, f(dx)
\end{aligned}$$

where to pass to the last line we exploited the fact that $|y - x| = d^{\partial\Omega}(x)$ for all $y \in \Pi_{\partial\Omega}(x)$, thus p^x-a.e. The last integral equals to zero since $[p^x] = x$ for every $x \in \operatorname{spt} f$. □

Remark 5.78. The sufficient condition in Proposition 5.77 can be proved to be the necessary condition as well. For that purpose, however, one requires a deep result of Smirnov (see [156]) in order to show that any solution of (FMD) problem is of the form λ_γ, see also [76]. This lies beyond the scope of this thesis and the reader is referred to [29] for the "if and only if version" of Proposition 5.77.

Remark 5.79. In the case when f is a signed measure Proposition 5.77 (and its stronger variant advertised in Remark 5.78 above) is no longer valid. It is enough to consider the example presented in Fig. 5.12(c) where the solution $\varsigma_{\mathscr{S}}$ may be proved to solve (up to a multiplicative constant) (FMD) in spite of the fact that $\operatorname{spt} f \not\subseteq M(\Omega)$. The proof of optimality for (FMD) may e.g. go through optimality conditions (4.42) where u can be constructed as in Example 5.76 (up to dividing by $\sqrt{2}$).

5.6.2. The Monge-Kantorovich cost function induced by a maximal monotone map

While assuming convexity of the design domain Ω, in this section we study the function $\ell_{\mathbf{v}} : \overline{\Omega} \times \overline{\Omega} \to \mathbb{R}_+$ that for a maximal monotone map $\mathbf{v} \in \mathbf{V}_\Omega$ and a pair of points $(a, b) \in \mathbb{R}^d \times \mathbb{R}^d$ is defined by:

$$\ell_{\mathbf{v}}(a,b) := \min \left\{ \sqrt{2\langle b' - a', b - a \rangle} \;:\; a' \in \mathbf{v}(a),\;\; b' \in \mathbf{v}(b) \right\}. \tag{5.122}$$

For the definition and properties of $\mathbf{V}_\Omega$ the reader is referred to Section 5.3.2, Proposition 5.29 in particular. Since for every $x \in \mathbb{R}^d$ the set $\mathbf{v}(x)$ is compact the minimum above is clearly attained. Next, as announced in introduction to Section 5.6, we shall construct the symmetric sub-additive envelope $c_{\mathbf{v}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ of $\ell_{\mathbf{v}}$. For each pair (a, b) it will be convenient to use a univariate maximal monotone map $\mathbf{v}_{a,b} : \mathbb{R} \to 2^{\mathbb{R}}$ defined as below

$$\mathbf{v}_{a,b}(t) := \big\langle \mathbf{v}\big((1-t)\,a + t\,b\big), b - a\big\rangle \qquad \forall\, t \in \mathbb{R} \tag{5.123}$$

(note the slight difference with respect to $\mathbf{v}_{a,b}$ defined in (5.59)). For each $t \in \mathbb{R}$ we shall additionally write $\mathbf{v}_{a,b}(t_-) = \min\big\{t' \in \mathbf{v}_{a,b}(t)\big\}$ and $\mathbf{v}_{a,b}(t_+) = \max\big\{t' \in \mathbf{v}_{a,b}(t)\big\}$ which may be termed the left sided and, respectively, the right sided limit of $\mathbf{v}_{a,b}$ at t.

The basic properties of the function $\ell_{\mathbf{v}}$ are summed up below:

Lemma 5.80. *Let* $\mathbf{v}$ *be an element of* $\mathbf{V}_\Omega$*. Then:*

(i) The function $\ell_{\mathbf{v}}$ *is lower semi-continuous on* $\mathbb{R}^d \times \mathbb{R}^d$*.*

(ii) Let $a, b \in \mathbb{R}^d$ *and* $\mathbf{v}_{a,b}$ *as defined in* (5.123)*. Then* $\mathbf{v}_{a,b}$ *is a maximal monotone map from* $\mathbb{R}$ *to* $\mathbb{R}$ *and*

$$\ell_{\mathbf{v}}(a, b) \;=\; \sqrt{2}\,\big(\mathbf{v}_{a,b}(1_-) - \mathbf{v}_{a,b}(0_+)\big)^{\frac{1}{2}}.$$

(iii) The following upper bounds hold true:

$$\ell_{\mathbf{v}}(a, b) \le \sqrt{2}\,|b-a|^{1/2}(D + |b-a|)^{1/2} \qquad \forall\,(a, b) \in \mathbb{R}^d \times \mathbb{R}^d, \tag{5.124}$$

$$\ell_{\mathbf{v}}(a, b) \le \sqrt{2D}\,|b-a|^{1/2} \qquad \text{if } (a, b) \in \overline{\Omega} \times \overline{\Omega}. \tag{5.125}$$

where D *is the diameter of* Ω*.*

Proof. The lower semi-continuity property stated in (i) is straightforward once we know that for any sequence $(a_n, b_n) \to (a, b)$ the elements $a'_n \in \mathbf{v}(a_n), b'_n \in \mathbf{v}(a_n)$ which realize the minimum in (5.122) remain in a compact subset of $\mathbb{R}^d$, thus (possibly after extraction of a subsequence) converge respectively to $a' \in \mathbf{v}(a)$ and $b' \in \mathbf{v}(b)$. It is in fact the case since, by the assertion (iii) of Proposition 5.29, one has $|a'_n - a_n| + |b'_n - b_n| \le 2D$. Then, owing to closedness of the graph of $\mathcal{G}_{\mathbf{v}}$, there is a subsequence of indices n_k such that $a'_{n_k} \to a'$, $b'_{n_k} \to b'$ where $a' \in \mathbf{v}(a)$, $b' \in \mathbf{v}(b)$ and moreover

$$\liminf_n \ell_{\mathbf{v}}(a_n, b_n) = \lim_k \ell_{\mathbf{v}}(a_{n_k}, b_{n_k}) = \lim_k \sqrt{2\langle b'_{n_k} - a'_{n_k}, b_{n_k} - a_{n_k}\rangle} = \sqrt{2\langle b' - a', b - a\rangle}.$$

The left hand side of the chain of equalities above is not smaller than $\ell_{\mathbf{v}}(a, b)$ by definition.

The equality in assertion (ii) is trivial since $\ell_{\mathbf{v}}(a, b) = \ell_{\mathbf{v}_{a,b}}(0, 1)$ while one has $\mathbf{v}_{a,b}(t) = [\mathbf{v}_{a,b}(t_-), \mathbf{v}_{a,b}(t_+)]$ for every $t \in \mathbb{R}$. On the other hand, if $a, b \in \Omega \times \Omega$ the intersection of the line $\big\{(1-t)a + tb : t \in \mathbb{R}\big\}$ with $\partial\Omega$ gives two values t_a, t_b such that $t_a < 0 < 1 < t_b$ and $(t_b - t_a)|b - a| \le D$. Therefore, since $\mathbf{v}$ agrees with the identity in $\mathbb{R}^d \setminus \overline{\Omega}$, we get

$$\mathbf{v}_{a,b}(1_-) - \mathbf{v}_{a,b}(0_+) \;\le\; \mathbf{v}_{a,b}(t_b - 0) - \mathbf{v}_{a,b}(t_a + 0) \;\le\; D\,|b - a|$$

hence we obtain the inequality $\ell_{\mathbf{v}}(a, b) \le \sqrt{2D}|b - a|^{1/2}$ for all $(a, b) \in \Omega \times \Omega$ which can be extended to $\overline{\Omega} \times \overline{\Omega}$ by using the lower semi-continuity of $\ell_{\mathbf{v}}$ obtained in assertion (i). Thus inequality (5.125) is proved. Inequality (5.124) follows if $(a, b) \in \overline{\Omega} \times \overline{\Omega}$ whereas it obviously holds true also for $(a, b) \in (\mathbb{R}^d \setminus \Omega)^2$. Eventually it is enough to check the case

where $a \in \Omega$ while $b \notin \overline{\Omega}$. This is done by considering the unique $b' \in \partial\Omega \cap [a,b]$ which is associated with a value $t_{b'} < 1$. Then

$$\mathbf{v}_{a,b}(1_-) - \mathbf{v}_{a,b}(0_+) \leq \big(\mathbf{v}_{a,b}(1_-) - \mathbf{v}_{a,b}(t_{b'}+0)\big) + \big(\mathbf{v}_{a,b}(t_{b'}-0) - \mathbf{v}_{a,b}(0_+)\big) \leq D|b-a| + |b-a|^2.$$

The proof of Lemma 5.80 is complete.

□

The next result concerns the behaviour of the map $\mathbf{V}_\Omega \ni \mathbf{v} \mapsto \ell_{\mathbf{v}}$ with $\mathbf{V}_\Omega = (\mathbf{V}_\Omega, \mathbf{h})$ being endowed with topology of graph convergence (cf. Section 5.3.2 for details):

Lemma 5.81. *Let $\{\mathbf{v}_n\}$ be a sequence in $\mathbf{V}_\Omega$ such that $\mathbf{v}_n \to \mathbf{v}$. Then $\ell_{\mathbf{v}_n} \to \ell_{\mathbf{v}}$ in the sense of Γ-convergence that is*

$$\begin{cases} \liminf_n \ell_{\mathbf{v}_n}(a_n, b_n) \geq \ell_{\mathbf{v}}(a,b) \quad \textit{whenever} \quad (a_n,b_n) \to (a,b) \textit{ in } \mathbb{R}^d \times \mathbb{R}^d, & \text{(5.126a)} \\ \forall (a,b) \quad \exists\, (a_n,b_n) \to (a,b) \quad \textit{such that} \quad \limsup_n \ell_{\mathbf{v}_n}(a_n,b_n) \leq \ell_{\mathbf{v}}(a,b). & \text{(5.126b)} \end{cases}$$

Proof. For proving (5.126a) we write $\ell_{\mathbf{v}_n}(a_n,b_n) = \sqrt{2\langle b'_n - a'_n, b_n - a_n\rangle}$ for suitable $a'_n \in \mathbf{v}(a_n)$, $b'_n \in \mathbf{v}(b_n)$. As a'_n, b'_n remain bounded (by assertion (iii) of Proposition 5.29) we may assume that, up to extraction of a subsequence, it holds that $a'_n \to a'$, $b'_n \to b'$. Then the graph convergence of $\mathbf{v}_n$ implies that $a' \in \mathbf{v}(a)$, $b' \in \mathbf{v}(b)$, hence $\liminf_n \ell_{\mathbf{v}_n}(a_n,b_n) = \sqrt{2\langle b'-a', b-a\rangle} \geq \ell_{\mathbf{v}}(a,b)$. On the other hand, for given a, b we can choose $a' \in \mathbf{v}(a)$, $b' \in \mathbf{v}(b)$ so that $\ell_{\mathbf{v}}(a,b) = \sqrt{2\langle b'-a', b-a\rangle}$. Then there exists $(a_n, a'_n) \to (a,a')$ and $(b_n, b'_n) \to (b,b')$ such that $a'_n \in \mathbf{v}_n(a_n)$, $b'_n \in \mathbf{v}_n(b_n)$. Then the property (5.126b) follows since

$$\limsup_n \ell_{\mathbf{v}_n}(a_n,b_n) \leq \limsup_n \sqrt{2\langle b'_n - a'_n, b_n - a_n\rangle} = \sqrt{2\langle b'-a', b-a\rangle} = \ell_{\mathbf{v}}(a,b),$$

which completes the proof.

□

Next we associate to every $\mathbf{v} \in \mathbf{V}_\Omega$ a cost function $c_{\mathbf{v}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ defined as follows:

$$c_{\mathbf{v}}(a,b) := \inf \left\{ \sum_{i=1}^{N-1} \ell_{\mathbf{v}}(x_i, x_{i+1}) \;:\; x_1 = a, \quad x_N = b, \quad N \geq 2 \right\}. \tag{5.127}$$

Note that the infimum in (5.127) is not reached in general since the number N of intermediate points is not upper bounded. Let us also remark that, by using the homeomorphism $\mathbf{i}$, we may associate the same cost to any $v \in \mathcal{V}_\Omega$ so that the notation c_v could be used as well.

From the very definition (5.127) we can infer that for every $u \in C(\mathbb{R}^d)$ there holds an equivalence:

$$u(y) - u(x) \;\leq\; \ell_{\mathbf{v}}(x,y) \quad \forall (x,y) \qquad \Leftrightarrow \qquad u(y) - u(x) \;\leq\; c_{\mathbf{v}}(x,y) \quad \forall (x,y). \tag{5.128}$$

The implication $\Leftarrow$ is simply because $c_{\mathbf{v}} \leq \ell_{\mathbf{v}}$ (take $N = 2$ in the definition (5.127)). The converse is due to the fact that for a finite sequence $\{x_i\}_{i=1}^N$ the sum $\sum_{i=1}^{N-1} \big(u(x_{i+1}) - u(x_i)\big)$ is not greater than $\sum_{i=1}^{N-1} \ell_{\mathbf{v}}(x_i, x_{i+1})$ and telescopes to $u(x_N) - u(x_1) = u(y) - u(x)$.

Theorem 5.82. *Let* $\mathbf{v}$ *be an element of* $\mathbf{V}_\Omega$*. Then:*

(i) $c_{\mathbf{v}}$ *is a continuous sub-additive function that for* $(a,b) \in \mathbb{R}^d \times \mathbb{R}^d$ *satisfies*

$$c_{\mathbf{v}}(a,b) \;\le\; \ell_{\mathbf{v}}(a,b) \;\le\; \sqrt{2}\,|b-a|^{1/2}(D+|b-a|)^{1/2}. \tag{5.129}$$

(ii) The following dual representation of the pseudo-metric $c_{\mathbf{v}}$ *holds true*

$$c_{\mathbf{v}}(a,b) = \max\Big\{u(b)-u(a)\,:\,u\in C(\mathbb{R}^d),\;\; u(y)-u(x)\le \ell_{\mathbf{v}}(x,y)\quad \forall (x,y)\Big\} \tag{5.130}$$

while the maximum is achieved. As a result $c_{\mathbf{v}}$ *is the largest symmetric and sub-additive minorant of the function* $\ell_{\mathbf{v}}$*.*

(iii) For every $(a,b)\in\mathbb{R}^d\times\mathbb{R}^d$ *the evaluation function* $(\mathbf{V}_\Omega,\mathbf{h})\ni \mathbf{v}\mapsto\; c_{\mathbf{v}}(a,b)$ *is concave and upper semi-continuous.*

Remark 5.83. The evaluation map in assertion (iii) above is not continuous in general. Indeed, consider for instance the case where $\Omega=(0,1)$ and $\mathbf{v}_n$ is the monotone map associated with the "staircase function" v_n on $[0,1]$ such that $v_n(0)=0,\;v_n(1)=1$ and $v_n' = \frac{1}{n-1}\sum_{i=1}^{n-1}\delta_{\frac{i}{n}}$. Then, by taking intermediate points in (5.127) very close to each i/n from the left and from the right, it is easy to check that $c_{\mathbf{v}_n}(0,1)=0$. Clearly $\mathbf{v}_n\to\mathbf{v}$ for $\mathbf{v}$ being the identity. Hence $0=\lim_n c_{\mathbf{v}_n}(0,1) < c_{\mathbf{v}}(0,1)=1$.

Proof. The sub-additivity of $c_{\mathbf{v}}$ is straightforward from the definition (5.127); the second inequality in (5.129) is obtained by taking $N=2$ and $(x_1,x_2)=(a,b)$ while the third one (already in (5.124)) implies that $c_{\mathbf{v}}$ is continuous on the diagonal hence everywhere by exploiting the sub-additivity property.

Let us prove (ii). The equivalence (5.128) implies that $c_{\mathbf{v}}(a,b)$ is not larger than the right hand side of (5.130). The converse inequality is obtained by considering the function $u(x)=c_{\mathbf{v}}(a,x)$ which satisfies $u(y)-u(x)\le c_{\mathbf{v}}(x,y)$ by sub-additivity and symmetry of $c_{\mathbf{v}}$. In addition, this function u is Hölder continuous by the assertion (i). This proves the equality in (5.130). Moreover, since the Hölder coefficient is uniformly bounded for all admissible u, the supremum is actually a maximum. The construction of u shows that there cannot be any symmetric and sub-additive minorant $\tilde{c}_{\mathbf{v}}$ of $\ell_{\mathbf{v}}$ for which $\tilde{c}_{\mathbf{v}}(a,b) > c_{\mathbf{v}}(a,b)$ for some pair $(a,b)\in\mathbb{R}^d\times\mathbb{R}^d$.

It remains to prove assertion (iii). Let us check first that that the map $\mathbf{v}\mapsto c_{\mathbf{v}}(a,b)$ is upper semi-continuous. Let $\mathbf{v}_n\to\mathbf{v}$ in $\mathbf{V}_\Omega$ and let us show that $\limsup_n c_{\mathbf{v}_n}(a,b)\le c_{\mathbf{v}}(a,b)$ for fixed $(a,b)\in(\mathbb{R}^d)^2$. Given a finite set $\{x_i\,:\,1\le i\le N\}$ such that $x_1=a\,,\;x_N=b$, we can choose, for each i, an approximating sequence $x_{i,n}\to x_i$ such that $\ell_{\mathbf{v}_n}(x_{i,n},x_{i+1,n})\to \ell_{\mathbf{v}}(x,y)$. This is indeed a consequence of the Γ-convergence of $\ell_{\mathbf{v}_n}$ to $\ell_{\mathbf{v}}$ proved in Lemma 5.81. In accordance with (5.127), for each n we may estimate $c_{\mathbf{v}_n}(a,b)$ from above by using the finite sequence of $N+2$ points $\{a,x_{1,n},\dots,x_{N,n},b\}$ and we are led to

$$\begin{aligned}\limsup_n c_{\mathbf{v}_n}(a,b) &\le \limsup_n \left\{\ell_{\mathbf{v}_n}(a,x_{1,n}) + \sum_{i=1}^{N-1}\ell_{\mathbf{v}_n}(x_{i,n},x_{i+1,n}) + \ell_{\mathbf{v}_n}(x_{N,n},b)\right\}\\ &= \sum_{i=1}^{N-1}\ell_{\mathbf{v}}(x_i,x_{i+1}),\end{aligned}$$

where to obtain the latter equality we additionally notice that $\lim_n \ell_{\mathbf{v}_n}(a, x_{1,n}) = \lim_n \ell_{\mathbf{v}_n}(x_{N,n}, b) = 0$ as a consequence of the estimates (5.124), (5.125). By minimizing the right hand sum with respect to x_i's we obtain the claimed upper semi-continuity inequality.

Let us now prove the concavity property. Thanks to the upper semi-continuity showed above it is enough to check the middle point property that is, for given $\mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}_\Omega$ and setting $\mathbf{v} = \frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2)$:

$$c_{\mathbf{v}}(a, b) \geq \frac{1}{2}\big(c_{\mathbf{v}_1}(a, b) + c_{\mathbf{v}_2}(a, b)\big). \tag{5.131}$$

In view of assertion (ii), there exist two elements $u_1, u_2 \in C(\mathbb{R}^d)$ such that for $i \in \{1, 2\}$

$$c_{\mathbf{v}_i}(a, b) = u_i(b) - u_i(a) \qquad \text{and} \qquad u_i(y) - u_i(x) \ \leq \ \ell_{\mathbf{v}_i}(x, y) \quad \forall\, (x, y).$$

Then (5.131) follows once we can show that $u = \frac{1}{2}(u_1 + u_2)$ satisfies the constraint $u(y) - u(x) \leq \ell_{\mathbf{v}}(x, y)$ for every $(x, y) \in (\mathbb{R}^d)^2$ as well. This is a consequence of the following concavity property of the map $\mathbf{v} \mapsto \ell^2_{\mathbf{v}}(x, y) := \big(\ell_{\mathbf{v}}(x, y)\big)^2$:

$$\ell^2_{\frac{\mathbf{v}_1+\mathbf{v}_2}{2}}(x, y) \ \geq \ \frac{1}{2}\left(\ell^2_{\mathbf{v}_1}(x, y) + \ell^2_{\mathbf{v}_2}(x, y)\right). \tag{5.132}$$

Indeed, since $|u_i(y) - u_i(x)| \leq \ell_{\mathbf{v_i}}(x, y)$, by straightforward computations we will get

$$|u(y)-u(x)|^2 \leq \frac{1}{4}\big(\ell_{\mathbf{v}_1}(x, y)+\ell_{\mathbf{v}_2}(x, y)\big)^2 \leq \frac{1}{2}\big(\ell^2_{\mathbf{v}_1}(x, y)+\ell^2_{\mathbf{v}_2}(x, y)\big) \leq \ell^2_{\frac{\mathbf{v}_1+\mathbf{v}_2}{2}}(x, y) = \ell^2_{\mathbf{v}}(x, y).$$

Let us proof the claim (5.132). Let $\hat{x} \in \frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2)(x)$ and $\hat{y} \in \frac{1}{2}(\mathbf{v}_1 + \mathbf{v}_2)(y)$. Then we have $\hat{x} = \frac{1}{2}(\hat{x}_1 + \hat{x}_2),\ \hat{y} = \frac{1}{2}(\hat{y}_1 + \hat{y}_2)$ for suitable $\hat{x}_i \in \mathbf{v}_i(x), \hat{y}_i \in \mathbf{v}_i(y)$. Therefore, by the definition of $\ell_{\mathbf{v}_i}(x, y)$, we infer that

$$2\,\langle \hat{y} - \hat{x}, y - x\rangle = \langle \hat{y}_1 - \hat{x}_1, y - x\rangle + \langle \hat{y}_2 - \hat{x}_2, y - x\rangle \geq \frac{1}{2}\big(\ell^2_{\mathbf{v}_1}(x, y) + \ell^2_{\mathbf{v}_2}(x, y)\big),$$

hence the claim by choosing $(\hat{x}, \hat{y})$ optimal for $\ell_{\frac{\mathbf{v}_1+\mathbf{v}_2}{2}}(x, y)$. The proof of Theorem 5.82 is now complete. □

The features of the function $c_{\mathbf{v}}$ make it possible to revise the idea (5.118) and eventually rewrite the problem $(\overline{\mathscr{P}})$ to a *geometrical form* that paves the way to Kantorovich-Rubinstein duality that will be performed in the next section:

$$\sup_{(u,\mathbf{v})\in C_0(\Omega)\times\mathbf{V}_\Omega} \left\{ \int u\, df \ : \ u(y) - u(x) \leq c_{\mathbf{v}}(x, y) \quad \forall\, (x, y) \in \overline{\Omega} \times \overline{\Omega} \right\} \tag{$\mathscr{P}_{\mathrm{geo}}$}$$

The precise relation to the problem $(\overline{\mathscr{P}})$ is as follows:

Proposition 5.84. *Solution of the problem $(\mathscr{P}_{\mathrm{geo}})$ exists and there holds an equality*

$$\mathcal{Z} = \max \overline{\mathscr{P}} = \max \mathscr{P}_{\mathrm{geo}}.$$

Moreover, a pair $(u, \mathbf{v}) \in C_0(\Omega) \times \mathbf{V}_\Omega$ solves $(\mathscr{P}_{\mathrm{geo}})$ if and only if (u, w) is optimal for $(\mathscr{P})$, where $w = \mathrm{id} - \mathbf{i}(\mathbf{v})$.

Proof. First we show the existence of an optimal pair $(u, \mathbf{v})$ for $(\mathscr{P}_{\mathrm{geo}})$. If $(u_n, \mathbf{v}_n)$ is a maximizing sequence, then $\{u_n\}$ is equicontinuous as a consequence of the uniform upper bound estimate in (5.129) hence relatively compact in $C_0(\Omega)$ (by Ascoli's theorem) while we recall that $\mathbf{V}_\Omega$ is a compact metric space (see Lemma 5.30). By exploiting the upper

semi-continuity property of the map $\mathbf{v} \in \mathbf{V}_\Omega \mapsto c_{\mathbf{v}}(x,y)$ holding true for all x, y (see assertion (iii) in Theorem 5.82), we see that any cluster point $(u, \mathbf{v})$ satisfies the inequality constraint $u(y) - u(x) \le c_{\mathbf{v}}(x,y)$ and therefore is a solution of $(\mathscr{P}_{\rm geo})$. The equivalence between the formulations $(\overline{\mathscr{P}})$ and $(\mathscr{P}_{\rm geo})$ can be deduced by combining Theorem 5.32, Remark 5.33 and (5.128). □

5.6.3. Maximal Monge-Kantorovich metrics and saddle point formulation

All along this subsection we will assume that Ω is convex and that the load $f = f_+ - f_-$ is a signed measure where $f_+, f_- \in \mathcal{M}^b_+(\Omega)$ is the Jordan decomposition in accordance with (2.4). It will be sometimes useful to consider alternative decompositions $f = \nu_1 - \nu_2$ where $\nu_1, \nu_2 \in \mathcal{M}^b_+(\Omega)$ are not necessarily singular with respect to each other (in contrast with f_+, f_-). Measures ν_1 and ν_2 will act as a source and target measures in the Monge-Kantorovich transport problem associated with the $c_{\mathbf{v}}$-cost. Note that we assume that $\nu_1(\partial\Omega) = \nu_2(\partial\Omega) = 0$ and thus $f(\partial\Omega) = 0$, which is not restrictive due to the Dirichlet condition on $\partial\Omega$, but it will allow to avoid some extra manipulations on $\partial\Omega$.

Let us fix an element $\mathbf{v} \in \mathbf{V}_\Omega$ and define

$$W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1, \nu_2) := \inf_{\nu_1^0, \nu_2^0 \in \mathcal{M}_+(\partial\Omega)} W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) \tag{5.133}$$

where $c_{\mathbf{v}}$ play the role of a cost function in the Monge-Kantorovich optimal transport problem for two measures $\mu_1, \mu_2 \in \mathcal{M}_+(\overline{\Omega})$ as below

$$W_{c_{\mathbf{v}}}(\mu_1, \mu_2) := \inf \left\{ \int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}(x,y)\, \gamma(dxdy) \ : \ \gamma \in \Gamma(\mu_1, \mu_2) \right\},$$

where $W_{c_{\mathbf{v}}}(\mu_1, \mu_2)$ may be called the *Wasserstein distance between μ_1 and μ_2 for the $c_{\mathbf{v}}$-cost*. Note that $W_{c_{\mathbf{v}}}$ and $W^{\partial\Omega}_{c_{\mathbf{v}}}$ are symmetric and that $W_{c_{\mathbf{v}}}(\mu_1, \mu_2) < +\infty$ if and only if $\int \mu_1 = \int \mu_2$. If so then by using the lower semi-continuity of the map $(\mu_1, \mu_2) \mapsto W_{c_{\mathbf{v}}}(\mu_1, \mu_2)$ (that may be proved by Kanotorich-Rubin theorem, see Theorem 4.59), one checks easily that the infimum in (5.133) is actually a minimum.

Owing to continuity and sub-additivity of the cost function $c_{\mathbf{v}}$ we may finally utilize the advertised Kantorovich-Rubinstein duality that generalizes Lemma 4.66 which was stated for the Euclidean cost only:

Lemma 5.85. *Let $\nu_1, \nu_2 \in \mathcal{M}^b_+(\Omega)$ and $\mathbf{v} \in \mathbf{V}_\Omega$. Then*

$$W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1, \nu_2) = \max_{u \in C_0(\Omega)} \left\{ \int u\, d(\nu_1 - \nu_2) \ : \ u(y) - u(x) \ \le \ c_{\mathbf{v}}(x,y) \quad \forall\, (x,y) \in \overline{\Omega} \times \overline{\Omega} \right\}. \tag{5.134}$$

As a consequence $W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1, \nu_2) = W^{\partial\Omega}_{c_{\mathbf{v}}}(f_+, f_-)$ whenever $\nu_1 - \nu_2 = f$. Furthermore, the infimum in (5.133) *is attained in the convex weakly-* compact subset*

$$\mathcal{M}_{\nu_1,\nu_2}(\partial\Omega) := \left\{ (\nu_1^0, \nu_2^0) \in \big(\mathcal{M}_+(\partial\Omega)\big)^2 \ : \ \begin{array}{l} \int(\nu_1 + \nu_1^0) = \int(\nu_2 + \nu_2^0), \\ \int \nu_2^0 \le \int \nu_1, \quad \int \nu_1^0 \le \int \nu_2 \end{array} \right\}.$$

Proof. Let $x_0 \in \partial\Omega$. By the uniform Hölder estimate (5.125), (5.129) and Ascoli's theorem, the set

$$\mathcal{B} := \Big\{ u \in C(\overline{\Omega}) \,:\, u(x_0) = 0, \quad u(y) - u(x) \le c_{\mathbf{v}}(x,y) \;\; \forall\, (x,y) \in \overline{\Omega} \times \overline{\Omega} \Big\}$$

is a convex compact subset of $C(\overline{\Omega})$. Since the cost function $c_{\mathbf{v}}$ is continuous and sub-additive, we know from Kantorovich-Rubinstein theorem (see Theorem 4.59 and Remark 4.60) that for every $(\nu_1^0, \nu_2^0) \in (\mathcal{M}_+(\partial\Omega))^2$, the following equality holds:

$$W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) \;=\; \sup_{u \in \mathcal{B}} \left\{ \int u \, d\big((\nu_1 + \nu_1^0) - (\nu_2 + \nu_2^0)\big) \right\}. \tag{5.135}$$

By applying Ky-Fan's result (Theorem 2.5) which allows switching the symbols inf and sup, we are led to

$$\begin{aligned} W_{c_{\mathbf{v}}}^{\partial\Omega}(\nu_1, \nu_2) \;&=\; \inf_{(\nu_1^0, \nu_2^0) \in (\mathcal{M}_+(\partial\Omega))^2} \; \sup_{u \in \mathcal{B}} \int u \, d\big((\nu_1 + \nu_1^0) - (\nu_2 + \nu_2^0)\big) \\ &= \sup_{u \in \mathcal{B}} \left\{ \int u \, d(\nu_1 - \nu_2) + \inf_{(\nu_1^0, \nu_2^0) \in (\mathcal{M}_+(\partial\Omega))^2} \int u \, d(\nu_1^0 - \nu_2^0) \right\} \end{aligned}$$

hence (5.134) by noticing that the infimum in the bottom line is finite if and only if $u = 0$ on $\partial\Omega$.

As a consequence of (5.134), we have $W_{c_{\mathbf{v}}}^{\partial\Omega}(\nu_1, \nu_2) = W_{c_{\mathbf{v}}}^{\partial\Omega}(f_+, f_-)$ whenever $\nu_1 - \nu_2 = f_+ - f_-$. It remains to show the last assertion of Lemma 5.85. As $\partial\Omega$ is compact, it can be readily checked that $\mathcal{M}_{\nu_1,\nu_2}(\partial\Omega)$ is a weakly-* convex compact subset of $(\mathcal{M}_+(\partial\Omega))^2$ thus, recalling (5.133), we have:

$$W_{c_{\mathbf{v}}}^{\partial\Omega}(\nu_1, \nu_2) \le \min \Big\{ W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) \;:\; (\nu_1^0, \nu_2^0) \in \mathcal{M}_{\nu_1,\nu_2}(\partial\Omega) \Big\}$$

where the existence of a minimizer on the right side is ensured by the lower semi-continuity of the function $(\nu_1^0, \nu_2^0) \mapsto W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0)$ (it is a point-wise supremum of a family of linear continuous functions, cf. (5.135)). Consequently, our argument boils down to showing that the inequality above is in fact an equality. This will be indeed the case if for every (ν_1^0, ν_2^0) we are able to construct measures $(\tilde{\nu}_1^0, \tilde{\nu}_2^0) \in \mathcal{M}_{\nu_1,\nu_2}(\partial\Omega)$ such that $W_{c_{\mathbf{v}}}(\nu_1 + \tilde{\nu}_1^0, \nu_2 + \tilde{\nu}_2^0) \le W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0)$. To show this claim, we may assume that $W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) < \infty$. Hence $\int(\nu_1 + \nu_1^0) = \int(\nu_2 + \nu_2^0)$ and there exists an element $\gamma \in \Gamma(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0)$ such that $W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) = \int c_{\mathbf{v}} \, d\gamma$. Set $\tilde{\gamma} = \gamma - \gamma_0$ where $\gamma_0 := \gamma \llcorner (\partial\Omega \times \partial\Omega)$. Clearly the marginals of $\tilde{\gamma}$, further denoted by $\tilde{\nu}_1$ and $\tilde{\nu}_2$, satisfy $\tilde{\nu}_1 \llcorner \Omega = \nu_1$ and $\tilde{\nu}_2 \llcorner \Omega = \nu_2$. Let $\tilde{\nu}_2^0 = \tilde{\nu}_2 \llcorner \partial\Omega$ and $\tilde{\nu}_1^0 = \tilde{\nu}_1 \llcorner \partial\Omega$. Then it holds that $\tilde{\nu}_1 = \nu_1 + \tilde{\nu}_1^0$ and $\tilde{\nu}_2 = \nu_2 + \tilde{\nu}_2^0$. Furthermore we have $\int \tilde{\nu}_1^0 = \tilde{\nu}_1(\partial\Omega) = \gamma(\partial\Omega \times \Omega) \le \int \nu_2$ while $\int \tilde{\nu}_2^0 = \tilde{\nu}_2(\partial\Omega) = \gamma(\Omega \times \partial\Omega) \le \int \nu_1$. It follows that $(\tilde{\nu}_1^0, \tilde{\nu}_2^0)$ belongs to $\mathcal{M}_{\nu_1,\nu_2}(\partial\Omega)$, hence our claim since:

$$W_{c_{\mathbf{v}}}(\nu_1 + \tilde{\nu}_1^0, \nu_2 + \tilde{\nu}_2^0) \le \int c_{\mathbf{v}} \, d\tilde{\gamma} \le \int c_{\mathbf{v}} \, d\gamma \le W_{c_{\mathbf{v}}}(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0).$$

The proof is complete. □

Based on the second part of the lemma above we deduce that for the set

$$\Gamma^{\partial\Omega}(\nu_1, \nu_2) := \Big\{ \gamma \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega}) \;:\; \exists (\nu_1^0, \nu_2^0) \in \mathcal{M}_{\nu_1,\nu_2}(\partial\Omega), \quad \gamma \in \Gamma(\nu_1 + \nu_1^0, \nu_2 + \nu_2^0) \Big\}.$$

there holds

$$W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1,\nu_2)=\inf\left\{\int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}(x,y)\,\gamma(dxdy)\ :\ \gamma\in\Gamma^{\partial\Omega}(\nu_1,\nu_2)\right\}\qquad(\mathrm{OTP}^{\partial\Omega}_{c_{\mathbf{v}}})$$

namely there is no gap between the infimum above and the infimum in (5.133). It is straightforward to check that the set $\Gamma^{\partial\Omega}(\nu_1,\nu_2)$ is compact and since $c_{\mathbf{v}}$ is continuous on $\overline{\Omega}\times\overline{\Omega}$ the *optimal transport problem with a fee free zone $\partial\Omega$ and with respect to the cost $c_{\mathbf{v}}$*, in short $(\mathrm{OTP}^{\partial\Omega}_{c_{\mathbf{v}}})$, is well posed for any $\mathbf{v}\in\mathbf{V}_\Omega$.

At last we can find that the number $\mathcal{Z}$ – that is equal to $\max\overline{\mathscr{P}}=\min\mathscr{P}^*=\inf\widetilde{\mathscr{P}^*}$ and that virtually characterizes the optimal membrane problem – can be recast as Monge-Kantorovich cost $W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1,\nu_2)$ that is maximal among all the maximal monotone maps $\mathbf{v}$ in the set $\mathbf{V}_\Omega$:

Theorem 5.86. *For any $\nu_1,\nu_2\in\mathcal{M}^b_+(\Omega)$ such that $\nu_1-\nu_2=f$ there holds*

$$\mathcal{Z}=\max\mathscr{P}_{\mathrm{geo}}=\max\left\{W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1,\nu_2)\ :\ \mathbf{v}\in\mathbf{V}_\Omega\right\}.\qquad(5.136)$$

In particular, for every pair $(u,\mathbf{v})$ solving $(\mathscr{P}_{\mathrm{geo}})$ the map $\mathbf{v}$ induces "the worst $c_{\mathbf{v}}$ metric", namely it solves the maximization problem on the right hand side.

Proof. The equality $\max\mathscr{P}_{\mathrm{geo}}=\sup\left\{W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1,\nu_2)\ :\ \mathbf{v}\in\mathbf{V}_\Omega\right\}$ follows from the formulation $(\mathscr{P}_{\mathrm{geo}})$ where, for every $\mathbf{v}\in\mathbf{V}_\Omega$, we compute the supremum with respect to u by applying (5.134). Clearly if $(\bar u,\bar{\mathbf{v}})$ solves $(\mathscr{P}_{\mathrm{geo}})$, then $\bar{\mathbf{v}}$ satisfies $W^{\partial\Omega}_{c_{\bar{\mathbf{v}}}}(\nu_1,\nu_2)\geq W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1,\nu_2)$ for all $\mathbf{v}\in\mathbf{V}_\Omega$. □

As the next step we can now make an almost one-to-one copy of the part of Section 4.8.2 in regarding the case $f\in\mathcal{M}^b_+(\Omega)$, i.e. when the load is non-negative (we do not enforce that f is a probability) and we may fix $\nu_1=f$, $\nu_2=0$. Then the problem $(\mathrm{OTP}^{\partial\Omega}_{c_{\mathbf{v}}})$ is about optimally transporting the mass f to the boundary – this time with respect to the $c_{\mathbf{v}}$-cost. Accordingly, $W^{\partial\Omega}_{c_{\mathbf{v}}}(f):=W^{\partial\Omega}_{c_{\mathbf{v}}}(f,0)$ may be termed *the Wasserstein distance of the mass f to the boundary $\partial\Omega$ for the $c_{\mathbf{v}}$-cost*. In this case it is easy to check that the condition $(\nu^0_1,\nu^0_2)\in\mathcal{M}_{f,0}(\partial\Omega)$ implies that $\nu^0_1=0$ and, as a result, $\gamma\in\mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$ is an element of $\Gamma^{\partial\Omega}(f,0)$ if and only if $\operatorname{spt}\gamma\subset\overline{\Omega}\times\partial\Omega$ and $\pi_1\#\gamma=f$. Then we can introduce notions that are analogous to those in Section 4.8.2 but with the euclidean distance $|y-x|$ substituted by $c_{\mathbf{v}}(x,y)$, namely:

$$d^{\partial\Omega}_{c_{\mathbf{v}}}(x):=\inf_{y\in\partial\Omega}c_{\mathbf{v}}(x,y),\qquad\Pi^{\partial\Omega}_{c_{\mathbf{v}}}(x):=\left\{y\in\partial\Omega\ :\ c_{\mathbf{v}}(x,y)=d^{\partial\Omega}_{c_{\mathbf{v}}}(x)\right\}$$

and

$$\mathcal{G}^{\partial\Omega}_{c_{\mathbf{v}}}:=\left\{(x,y)\in\overline{\Omega}\times\partial\Omega\ :\ y\in\Pi^{\partial\Omega}_{c_{\mathbf{v}}}(x)\right\}.$$

Let us recall that by virtue of Theorem 5.82 the function $c_{\mathbf{v}}:\overline{\Omega}\times\overline{\Omega}\to\mathbb{R}_+$ is symmetric, subadditive (it is a pseudo-distance) and continuous. Accordingly, by repeating the arguments of proof of Proposition 4.68 we deduce that $d^{\partial\Omega}_{c_{\mathbf{v}}}(x)-d^{\partial\Omega}_{c_{\mathbf{v}}}(y)\leq c_{\mathbf{v}}(x,y)$ for every $(x,y)\in\overline{\Omega}\times\overline{\Omega}$. Since in addition $c_{\mathbf{v}}(x,y)\leq\ell_{\mathbf{v}}(x,y)\leq\sqrt{2\operatorname{diam}\Omega}\,|y-x|^{\frac12}$ we can see that $d^{\partial\Omega}_{c_{\mathbf{v}}}:\overline{\Omega}\to\mathbb{R}$ is Hölder continuous while it clearly vanishes on $\partial\Omega$. Thus $\mathcal{G}^{\partial\Omega}_{c_{\mathbf{v}}}$ is a closed subset of $\overline{\Omega}\times\partial\Omega$ and the "$c_{\mathbf{v}}$-counterpart" of Proposition 4.68 may be readily given.

Proposition 5.87. *Let us consider a non-negative load $f \in \mathcal{M}^b_+(\Omega)$. Then a transportation plan $\gamma \in \mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$ solves problem* $(\mathrm{OTP}^{\partial\Omega}_{c_\mathbf{v}})$ *for $\nu_1 = f$, $\nu_2 = 0$ if and only if the following conditions are met:*

(i) $\pi_1\#\gamma = f$,
(ii) $\operatorname{spt}\gamma \subset \mathcal{G}^{\partial\Omega}_{c_\mathbf{v}}$,

while the Wasserstein distance of f to boundary $\partial\Omega$ equals

$$W^{\partial\Omega}_{c_\mathbf{v}}(f) = \int_{\overline{\Omega}} d^{\partial\Omega}_{c_\mathbf{v}}(x)\, f(dx).$$

Moreover, if $\mathbf{v}$ solves (5.136) *then $W^{\partial\Omega}_{c_\mathbf{v}}(f) = \max\mathcal{P}_{\mathrm{geo}} = \mathcal{Z}$ while for $u = d^{\partial\Omega}_{c_\mathbf{v}}$ the pair $(u,\mathbf{v})$ solves* $(\mathscr{P}_{\mathrm{geo}})$.

Proof. The proof of the main part of the statement is fully analogous to the proof of Proposition 4.68 (see also Remark 4.69), including optimality of $u = d^{\partial\Omega}_{c_\mathbf{v}}$ in the problem (5.136) for a fixed $\mathbf{v} \in \mathbf{V}_\Omega$. The "moreover part" is then clear when acknowledging Theorem 5.86. □

Clearly $d^{\partial\Omega}_{c_\mathbf{v}}$ is a non-negative function therefore, thanks to Proposition 5.84, we immediately infer that:

Corollary 5.88. *For a non-negative load $f \in \mathcal{M}^b_+(\Omega)$ there always exists a non-negative function $u \in C_0(\Omega;\mathbb{R}_+)$ that together with some $w \in \mathcal{W}_\Omega$ solves* $(\overline{\mathscr{P}})$.

In general solution of $(u,\mathbf{v})$ is non-unique therefore the relation $u = d^{\partial\Omega}_{c_\mathbf{v}}$ does not have to hold (when $f \geq 0$). We may, however show that it is true at points where the loads are applied, namely:

Corollary 5.89. *For a non-negative load $f \in \mathcal{M}^b_+(\Omega)$ every solution $(u,\mathbf{v})$ of* $(\mathscr{P}_{\mathrm{geo}})$ *satisfies*

$$u(x) = d^{\partial\Omega}_{c_\mathbf{v}}(x) \qquad \forall\, x \in \operatorname{spt} f. \tag{5.137}$$

Proof. First, we observe that function u satisfies $u(x) - u(y) \leq c_\mathbf{v}(x,y)$ for each $(x,y) \in \overline{\Omega}\times\overline{\Omega} \subset \overline{\Omega}\times\partial\Omega$ therefore there must hold: $u(x) \leq d^{\partial\Omega}_{c_\mathbf{v}}(x)$ at every point $x \in \overline{\Omega}$. Assume that there exists $x_0 \in \operatorname{spt} f$ such that $u(x_0) \leq d^{\partial\Omega}_{c_\mathbf{v}}(x_0) - 2\varepsilon$ for some $\varepsilon > 0$. Then, by continuity of both u and $d^{\partial\Omega}_{c_\mathbf{v}}$ we have $u(x) \leq d^{\partial\Omega}_{c_\mathbf{v}}(x) - \varepsilon$ for every $x \in \mathscr{U} := B(x_0,\delta)$ for some $\delta > 0$. As a result

$$\int_{\overline{\Omega}} u\, df = \int_{\mathscr{U}} u\, df + \int_{\overline{\Omega}\setminus\mathscr{U}} u\, df \leq \int_{\mathscr{U}} \left(d^{\partial\Omega}_{c_\mathbf{v}} - \varepsilon\right) df + \int_{\overline{\Omega}\setminus\mathscr{U}} d^{\partial\Omega}_{c_\mathbf{v}}\, df = -\varepsilon\, f(\mathscr{U}) + W^{\partial\Omega}_{c_\mathbf{v}}(f)$$

while $W^{\partial\Omega}_{c_\mathbf{v}}(f) = \max\mathcal{P}_{\mathrm{geo}}$ by Proposition 5.87. Since $\mathscr{U}$ is a neighbourhood of a point in $\operatorname{spt} f$ there must hold $f(\mathscr{U}) > 0$. This contradicts the pair $(u,\mathbf{v})$ being a solution of $(\mathscr{P}_{\mathrm{geo}})$. □

Remark 5.90. As a consequence of Corollary 5.89 we can be sure that functions u (or approximations u_h) showed in Fig. 5.2(a), 5.9(b), 5.11(b) – that is for examples in which $\operatorname{spt} f = \overline{\Omega}$ – are equal to $d^{\partial\Omega}_{c_\mathbf{v}}$ for $\mathbf{v} = \mathrm{id} - \mathbf{i}^{-1}(w)$ where w (or w_h resp.) is the corresponding solution that was found. In particular the function in Fig. 5.9(b) is the $c_\mathbf{v}$-counterpart of the sandpile solution showed in Fig. 4.11(a) which was computed for $c(x,y) = |y-x|$.

Let us consider a solution $\overline{\mathbf{v}} \in \mathbf{V}_\Omega$ of the problem (5.136) (or equivalently $\overline{\mathbf{v}}$ that together with some $\overline{u}$ solve $(\mathscr{P}_{\rm geo})$). Then it is natural to take a solution $\overline{\gamma}$ of the problem $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ and ask what is the connection between $\overline{\gamma}$ and optimal fields involved in the membrane problem. e.g. pairs $(\overline{\vartheta}, \overline{\varsigma})$ solving $(\mathscr{P}^*)$ or pairs $(\overline{\pi}, \overline{\mathscr{S}})$ solving $(\widetilde{\mathscr{P}^*})$. This issue proves to be very subtle and will be discussed in more details in Section 5.7.4. We know that in general solution of optimal transport problem is non-unique – it will turn out that optimizing the cost $c_{\mathbf{v}}$ will "increase the level of the non-uniqueness". In the next section it will appear that more than optimality of $\overline{\gamma}$ for $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ is required to make a connection to the optimal design of membranes: we need that the pair $(\overline{\gamma}, \overline{\mathbf{v}})$ is a saddle point for the Lagrangian

$$\Gamma^{\partial\Omega}(\nu_1, \nu_2) \times \mathbf{V}_\Omega \ni (\gamma, \mathbf{v}) \quad \mapsto \quad \int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}\, d\gamma. \tag{5.138}$$

Below we show that such a saddle point always exists:

Theorem 5.91. *For every $\nu_1, \nu_2 \in \mathcal{M}^b_+(\Omega)$ such that $\nu_1 - \nu_2 = f$ there exists a pair $(\overline{\gamma}, \overline{\mathbf{v}}) \in \Gamma^{\partial\Omega}(\nu_1, \nu_2) \times \mathbf{V}_\Omega$ solving the saddle point problem:*

$$\int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}\, d\overline{\gamma} \;\le\; \int_{\overline{\Omega}\times\overline{\Omega}} c_{\overline{\mathbf{v}}}\, d\overline{\gamma} \;\le\; \int_{\overline{\Omega}\times\overline{\Omega}} c_{\overline{\mathbf{v}}}\, d\gamma \qquad \forall\, \gamma \in \Gamma^{\partial\Omega}(\nu_1, \nu_2), \quad \forall\, \mathbf{v} \in \mathbf{V}_\Omega. \tag{5.139}$$

Furthermore, for any such an optimal pair $(\overline{\gamma}, \overline{\mathbf{v}})$, there exists an optimal potential $\overline{u} \in C_0(\Omega)$ such that $(\overline{u}, \overline{\mathbf{v}})$ solves the geometric problem $(\mathscr{P}_{\rm geo})$ and satisfies the relations

$$\overline{u}(x) - \overline{u}(y) \le c_{\overline{\mathbf{v}}}(x, y) \quad \forall\, (x, y) \in \overline{\Omega}^2, \qquad \overline{u}(x) - \overline{u}(y) = c_{\overline{\mathbf{v}}}(x, y) \quad \text{for } \overline{\gamma}\text{-a.e. } (x, y) \in \overline{\Omega}^2. \tag{5.140}$$

Remark 5.92. By applying Theorem 5.91 with $\nu_1 = f_+$ and $\nu_2 = f_-$, we see that a pair $(\overline{u}, \overline{\mathbf{v}})$ is a solution of the geometric problem $(\mathscr{P}_{\rm geo})$ if and only if there exists a suitable $\overline{\gamma}$ such that the triple $(\overline{\gamma}, \overline{\mathbf{v}}, \overline{u})$ belongs to $\Gamma^{\partial\Omega}(f_+, f_-) \times \mathbf{V}_\Omega \times C_0(\Omega)$ and satisfies the two conditions (5.139) and (5.140).

Proof. As can be readily checked $\Gamma^{\partial\Omega}(\nu_1, \nu_2)$ is a convex compact subset of $\mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ endowed with the weak-* topology. Besides, from Lemma 5.81 we know that the convex set $\mathbf{V}_\Omega$ equipped with the distance $\mathbf{h}$ is compact. The existence of a saddle point for (5.139) then follows from the second assertion of Theorem 2.5. Indeed, the Lagrangian $L(\gamma, \mathbf{v}) := \int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}\, d\gamma$ is convex lower semi-continuous with respect to $\gamma \in \Gamma^{\partial\Omega}(\nu_1, \nu_2)$ and, as a consequence of combining assertion (iii) of Theorem 5.82 with Fatou's lemma, it is concave upper semi-continuous with respect to $\mathbf{v} \in \mathbf{V}_\Omega$.

Let now $(\overline{\gamma}, \overline{\mathbf{v}})$ satisfy (5.139). Then, the left hand side inequality in (5.139) implies that $W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1, \nu_2) \le \int c_{\overline{\mathbf{v}}}\, d\overline{\gamma}$ while the right hand side implies that $\int c_{\overline{\mathbf{v}}}\, d\overline{\gamma} = W^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(\nu_1, \nu_2)$. Therefore $\overline{\mathbf{v}}$ is a maximizer on the right hand side of (5.136) and we obtain the equalities

$$\mathcal{Z} = \max \mathscr{P}_{\rm geo} = W^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(\nu_1, \nu_2) = \int_{\overline{\Omega}\times\overline{\Omega}} c_{\overline{\mathbf{v}}}(x, y)\, \overline{\gamma}(dxdy). \tag{5.141}$$

Next, by applying Lemma 5.85 with $\mathbf{v} = \overline{\mathbf{v}}$, there exists an optimal potential $\overline{u}$ for (5.134). Clearly the equalities in (5.141) imply that $(\overline{u}, \overline{\mathbf{v}})$ solves $(\mathscr{P}_{\rm geo})$ while we have

$$\int_{\overline{\Omega}\times\overline{\Omega}} c_{\overline{\mathbf{v}}}(x, y)\, \overline{\gamma}(dxdy) = \int \overline{u}\, df = \int \overline{u}\, d(\nu_1 - \nu_2) = \int_{\overline{\Omega}\times\overline{\Omega}} \big(\overline{u}(x) - \overline{u}(y)\big)\, \overline{\gamma}(dxdy).$$

where in the last equality we used the fact that $\overline{u} = 0$ on $\overline{\Omega}$ and that $(\pi_i \# \gamma) \llcorner \Omega = \nu_i$ for $i = 1, 2$. Since $\overline{u}$ satisfies $\overline{u}(x) - \overline{u}(y) \le c_{\overline{\mathbf{v}}}(x, y)$ for all (x, y), the equality between the left hand side and the right hand side of the chain above is possible if and only if $\overline{u}$ satisfies the equality condition in (5.140). The last assertion of Theorem 5.91 follows. □

5.6.4. Optimal string systems revisited – a connection to the saddle point problem

One of the purposes of this subsection is to give a feeling of the role of optimal transport problem $\overline{\gamma}$ that together with a maximal monotone map $\overline{\mathbf{v}} \in \mathbf{V}_\Omega$ constitutes a saddle point of Lagrangian (5.138). This issue becomes much clearer when for given data Ω, f there exists an optimal string system, i.e. a solution $(\overline{\pi}, \overline{\mathscr{S}})$ to the problem $(\widetilde{\mathscr{P}^*})$.

To make a rigorous passage between the two problems, however, it will be necessary to bring the optimality conditions in Proposition 5.57 to fully general framework, i.e. when for (u, w) solving $(\overline{\mathscr{P}})$ the function w is discontinuous or, equivalently, for $(u, \mathbf{v})$ solving the geometric problem $(\mathscr{P}_{\mathrm{geo}})$ the map $\mathbf{v}$ is truly multivalued. This proves to be an intriguing mathematical challenge on its own. Moreover such general optimality conditions for string systems will be crucial in Chapter 6.

Let us start by defining for $\mathbf{v} \in \mathbf{V}_\Omega$ a function $\mathscr{L}_{\mathbf{v}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ that differs from $\ell_{\mathbf{v}}$ defined in (5.122) by the max sign in the place of the min sign:

$$\mathscr{L}_{\mathbf{v}}(a, b) := \max \Big\{ \sqrt{2\langle b' - a', b - a\rangle} \ : \ a' \in \mathbf{v}(a),\ b' \in \mathbf{v}(b) \Big\}. \tag{5.142}$$

Clearly $\mathscr{L}_{\mathbf{v}}(x, y) = \ell_{\mathbf{v}}(x, y)$ at pairs (x, y) where $\mathbf{v}$ is univalued but not only in such case: it is enough that the tangent components $\langle \mathbf{v}(x), y - x\rangle$ and $\langle \mathbf{v}(y), y - x\rangle$ are singletons. Moreover, for a boundary point $a \in \partial\Omega$, we may assign $a' = a$ in the maximum (5.142), namely:

Lemma 5.93. *The formula for $\mathscr{L}_{\mathbf{v}}$ simplifies for the points on the boundary:*

$$\frac{1}{2}\big(\mathscr{L}_{\mathbf{v}}(a, b)\big)^2 = \begin{cases} \max\big\{\langle b' - a, b - a\rangle : \ b' \in \mathbf{v}(b)\big\} & \text{if } (a, b) \in \partial\Omega \times \Omega, \\ |b - a|^2 & \text{if } (a, b) \in \partial\Omega \times \partial\Omega. \end{cases} \tag{5.143}$$

Proof. Owing to assertion (iii) of Proposition 5.29 we have $\mathbf{v}(x) = \{x\}$ if $x \notin \overline{\Omega}$ while $\mathbf{v}(x) \supset \{x\}$ if $x \in \partial\Omega$. The inequality $\ge$ in the claim is therefore trivial. Next, by convexity of Ω we have $a_n = a + \frac{1}{n}(a - b) \notin \Omega$, thus $\mathbf{v}(a_n) \supset \{a_n\}$ for every $n \in \mathbb{N}_+$. By the monotonicity property we deduce that, for every $a' \in \mathbf{v}(a)$, $0 \le n\langle a_n - a', a_n - a\rangle = \frac{1}{n}|a - b|^2 + \langle a - a', a - b\rangle$ hence $\langle a - a', b - a\rangle \le 0$ by sending $n \to \infty$. Accordingly we get $\langle b' - a', b - a\rangle \le \langle b' - a, b - a\rangle$ for every $a' \in \mathbf{v}(a)$ and $b' \in \mathbf{v}(b)$. If in addition $b \in \partial\Omega$, we may choose $b' = b$ on the right hand side of the latter inequality. This proves the desired upper bound for $\mathscr{L}_{\mathbf{v}}(a, b)$, hence (5.143). □

By studying the proof of the optimality conditions in Proposition 5.57 we may find that testing the equilibrium equation (ii) in (5.89) against $\phi = w$, where w was a candidate for solving $(\overline{\mathscr{P}})$, played a key part. This is *a priori* possible only if w is continuous. The

technical result below is the way to go around this equilibrium equation once the map $\mathbf{v} = \mathrm{id} - \mathbf{i}^{-1}(w)$ is multivalued:

Lemma 5.94. *Let us be given $\mathbf{v} \in \mathbf{V}_\Omega$ and $\mathscr{S} \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ satisfying condition (ii) in (5.89). Then we have*

$$\frac{1}{2} \int_{\overline{\Omega}\times\overline{\Omega}} \frac{\ell^2_{\mathbf{v}}(x,y)}{|y-x|}\, \mathscr{S}(dxdy) \le \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\, \mathscr{S}(dxdy), \tag{5.144}$$

with an equality if only if $\ell_{\mathbf{v}} = \mathscr{L}_{\mathbf{v}}$ holds true $\mathscr{S}$-a.e.

Proof. Assume first that $\mathbf{v}$ is univalued, i.e. $\mathbf{v}(x) = \{v(x)\}$. Then, since the graph $\mathcal{G}_{\mathbf{v}}$ is closed, there exists $w \in C_0(\Omega;\mathbb{R}^d)$ such that $v(x) = x - w(x)$ and the assumption on $\mathscr{S}$ implies that

$$\int_{\overline{\Omega}\times\overline{\Omega}} \Big\langle v(y) - v(x), \frac{y-x}{|y-x|} \Big\rangle\, \mathscr{S}(dxdy) = \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\, \mathscr{S}(dxdy), \tag{5.145}$$

which, in view of (5.122), means that (5.144) holds with an equality. Consider now a general $\mathbf{v} \in \mathbf{V}_\Omega$. We observe that, if we let η be the sum of the marginals of $\mathscr{S}$, the equality (5.145) can be extended by density to all elements v in the closure of $C_0(\Omega;\mathbb{R}^d)$ in $L^1_\eta(\overline{\Omega};\mathbb{R}^d)$, in particular to any bounded Borel function $v : \overline{\Omega} \to \mathbb{R}^d$ such that $v = \mathrm{id}$ η-a.e. in $\partial\Omega$. The idea consists now in choosing for v a suitable Borel measurable selection of the multifunction $x \mapsto \mathbf{v}(x)$. More precisely, let $v_0 : \overline{\Omega} \to \mathbb{R}^d$ be defined by:

$$v_0(x) = s_d\big(\mathbf{v}(x)\big) \quad \text{if } x \in \Omega, \qquad v_0(x) = x \quad \text{if } x \in \partial\Omega$$

where, for every convex compact subset $A \subset \mathbb{R}^d$, $s_d(A)$ stands for the Steiner centre given by

$$s_d(A) := \frac{1}{\omega_d} \int_{S^{d-1}} z\, \chi^*_A(z)\, \mathcal{H}^{d-1}(dz)$$

where ω_d denotes the volume of the unit ball in $\mathbb{R}^d$. The important feature of Steiner centre is that $s_d(A)$ belongs to $\mathrm{ri}(A)$, that is to the relative interior of A (note that this includes the case of a singleton $A = \{a\}$ when $s_d(A) = a$), see [155, page 50] and [151]. It is also well established that the map $2^{\mathbb{R}^d} \ni A \mapsto s_d(A) \in \mathbb{R}^d$ is Lipschitz continuous with respect to the Hausdorff distance on $2^{\mathbb{R}^d}$, cf. [155]. We recall that $\mathbf{v}$ is an upper semi-continuous multifunction, therefore by Theorem 2.37 it is also a Borel measurable multifunction, which by Theorem 2.35 amounts to saying that $\mathbb{R}^d \ni x \mapsto \mathbf{v}(x) \in 2^{\mathbb{R}^d}$ is Borel measurable for the Hausdorff distance on $2^{\mathbb{R}^d}$. Eventually $\mathbb{R}^d \ni x \mapsto s_d\big(\mathbf{v}(x)\big) \in \mathbb{R}^d$ must be Borel measurable as the composition of the two aforementioned mappings. It follows that the function v_0 defined above is Borel and satisfies $v_0(x) = x$ η-a.e. on $\partial\Omega$ and $v_0(x) \in \mathrm{ri}(\mathbf{v})(x)$ η-a.e. in Ω. Therefore v_0 satisfies the equality (5.145) while $\mathscr{S}$-a.e. $\frac{1}{2}\ell^2_{\mathbf{v}}(x,y) \le \langle v_0(y) - v_0(x), y - x\rangle \le \frac{1}{2}\mathscr{L}^2_{\mathbf{v}}(x,y)$. By integrating with respect to $\mathscr{S}$ we are led to

$$\frac{1}{2} \int_{\overline{\Omega}\times\overline{\Omega}} \frac{\ell^2_{\mathbf{v}}(x,y)}{|y-x|}\, \mathscr{S}(dxdy) \le \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\, \mathscr{S}(dxdy) \le \frac{1}{2} \int_{\overline{\Omega}\times\overline{\Omega}} \frac{\mathscr{L}^2_{\mathbf{v}}(x,y)}{|y-x|}\, \mathscr{S}(dxdy).$$

The inequality (5.144) follows and it becomes an equality provided that $\ell_{\mathbf{v}} = \mathscr{L}_{\mathbf{v}}$ holds $\mathscr{S}$-a.e. Conversely, let us assume that in (5.144) we have equality. Then, since the equality (5.145) holds for v_0, for $\mathscr{S}$-a.e. (x, y) we obtain:

$$\langle v_0(y) - v_0(x), y - x\rangle = \frac{1}{2}\,\ell^2_{\mathbf{v}}(x, y) = \min\Big\{\langle y' - x', y - x\rangle \;:\; x' \in \mathbf{v}(x),\; y' \in \mathbf{v}(y)\Big\}. \tag{5.146}$$

This means that the minimum of the linear form $(x', y') \mapsto \langle y' - x', y - x\rangle$ on the convex compact subset $\mathbf{v}(x)\times\mathbf{v}(y)$ is reached at $\big(v_0(x), v_0(y)\big)$. In particular, for $\mathscr{S}$-a.e. $(x, y) \in \Omega^2$, we will have $\big(v_0(x), v_0(y)\big) \in \mathrm{ri}(\mathbf{v})(x) \times \mathrm{ri}(\mathbf{v})(y)$ and this is not possible unless the linear form remains constant on $\mathbf{v}(x) \times \mathbf{v}(y)$. We thereby deduce that $\ell_{\mathbf{v}} = \mathscr{L}_{\mathbf{v}}$ is satisfied $\mathscr{S}$-a.e on $\Omega \times \Omega$. Let us now consider a pair $(x, y) \in \partial\Omega \times \Omega$ that satisfies (5.146). By prescribing $x' = x$ in the infimum defining $\ell_{\mathbf{v}}$ we get

$$\frac{1}{2}\,\ell^2_{\mathbf{v}}(x, y) = \langle v_0(y) - x, y - x\rangle \geq \min\Big\{\langle y' - x, y - x\rangle \;:\; y' \in \mathbf{v}(y)\Big\} \geq \frac{1}{2}\,\ell^2_{\mathbf{v}}(x, y),$$

which proves that each inequality is an equality and, as a result, the affine function $\mathbf{v}(y) \ni y' \mapsto \langle y' - x, y - x\rangle$ is minimal at $v_0(y) \in \mathrm{ri}(\mathbf{v})(y)$ thus it must be constant. It follows that

$$\frac{1}{2}\,\ell^2_{\mathbf{v}}(x, y) = \max\Big\{\langle y' - x, y - x\rangle \;:\; y' \in \mathbf{v}(y)\Big\} = \frac{1}{2}\,\mathscr{L}^2_{\mathbf{v}}(x, y),$$

where in the second equality we used (5.143). If $(x, y) \in \partial\Omega \times \partial\Omega$, (5.143) provides the equality $\frac{1}{2}\,\mathscr{L}^2_{\mathbf{v}}(x, y) = |y - x|^2$ while, owing to the first equality in (5.146) and the fact that $v_0(x) = x$ on $\partial\Omega$, we have $\frac{1}{2}\,\ell^2_{\mathbf{v}}(x, y) = |y - x|^2$. □

Theorem 5.95. *Assuming a convex domain Ω let $(\pi, \mathscr{S})$ be an element of $\mathcal{M}(\overline{\Omega}\times\overline{\Omega}; \mathbb{R}\times\mathbb{R}_+)$ given in the form $(\pi, \mathscr{S}) = (\alpha\mathscr{S}, \mathscr{S})$ with $\alpha \in L^1_{\mathscr{S}}(\overline{\Omega} \times \overline{\Omega})$ and let $(u, \mathbf{v}) \in C_0(\Omega) \times \mathbf{V}_\Omega$.*

Then the pairs $(u, \mathbf{v})$ and $(\pi, \mathscr{S})$ are optimal for, respectively, $(\mathscr{P}_{\mathrm{geo}})$ and $(\widetilde{\mathscr{P}^*})$ if and only the following conditions are satisfied:

$$\begin{cases} (i) & (\alpha\mathscr{S}, \mathscr{S}) \in \Xi(\Omega, f), \\ (ii) & u(y) - u(x) \leq \ell_{\mathbf{v}}(x, y) \qquad \forall (x, y) \in \overline{\Omega} \times \overline{\Omega}, \\ (iii) & |u(y) - u(x)| = \ell_{\mathbf{v}}(x, y) \qquad \textit{for } \mathscr{S}\textit{-a.e. } (x, y), \\ (iv) & \alpha(x, y) = \frac{u(y)-u(x)}{|y-x|} \qquad \textit{for } \mathscr{S}\textit{-a.e. } (x, y), \\ (v) & \ell_{\mathbf{v}}(x, y) = \mathscr{L}_{\mathbf{v}}(x, y) \qquad \textit{for } \mathscr{S}\textit{-a.e. } (x, y). \end{cases} \tag{5.147}$$

Moreover, if the above conditions are met, then for $\mathscr{S}$-a.e. (x, y) we have:

(1) $\ell_{\mathbf{v}}(x, y) = c_{\mathbf{v}}(x, y)$,
(2) u is affine on $[x, y]$,
(3) $\langle \mathbf{v}_{x,y}, \tau^{x,y}\rangle$ is univalued on $[x, y]$ and it is affine there.

Remark 5.96. The system of optimality criteria (5.147) above may be converted back to the format of Proposition 5.57, i.e. when solution (u, w) of $(\overline{\mathscr{P}})$ is considered. Then, for $\mathbf{w} = \mathbf{i}^{-1}(w)$ conditions (ii), (iii), (v) may be rewritten as

$$(ii') \qquad \frac{1}{2}\,|u(y)-u(x)|^2 + \langle y'-x', y-x\rangle \le |y-x|^2 \qquad \begin{array}{l}\forall\,(x,y)\in\overline{\Omega}\times\overline{\Omega},\\ \forall\,x'\in\mathbf{w}(x),\ \forall\,y'\in\mathbf{w}(y),\end{array}$$

$$(iii', v') \qquad \frac{1}{2}\,|u(y)-u(x)|^2 + \langle y'-x', y-x\rangle = |y-x|^2 \qquad \begin{array}{l}\text{for } \mathscr{S}\text{-a.e. } (x,y)\in\overline{\Omega}\times\overline{\Omega},\\ \forall\,x'\in\mathbf{w}(x),\ \forall\,y'\in\mathbf{w}(y).\end{array}$$

Clearly (ii) and (ii') are equivalent. While acknowledging (ii') we can infer from (iii', v') that $\langle y'-x', y-x\rangle$ is independent of the choice of $x' \in \mathbf{w}(x) = x - \mathbf{v}(x)$ and $y' \in \mathbf{w}(y) = y - \mathbf{v}(y)$ hence (v) is indeed encompassed by the condition (iii', v').

Proof. Conditions (i) and (ii) are equivalent to the admissibility of $(\pi, \mathscr{S})$ and $(u, \mathbf{v})$ for $(\widetilde{\mathscr{P}^*})$ and $(\mathscr{P}_{\rm geo})$ respectively. Therefore they are assumed to hold true along the proof. By using (ii) and Young's inequality, we derive the following inequalities

$$\begin{aligned}\alpha(x,y)\big(u(y)-u(x)\big) \le |\alpha(x,y)||u(y)-u(x)| &\le |\alpha(x,y)|\,\ell_{\mathbf{v}}(x,y) \\ &\le \frac{1}{2}\,|y-x|\,\alpha^2(x,y) + \frac{1}{2}\,\frac{\ell^2_{\mathbf{v}}(x,y)}{|y-x|},\end{aligned} \tag{5.148}$$

which clearly become equalities if and only if conditions (iii) and (iv) are satisfied. By integrating with respect to $\mathscr{S}$ and thanks to (5.144) we infer that

$$\begin{aligned}\int u\,df &= \int_{\overline{\Omega}\times\overline{\Omega}} \alpha(x,y)\big(u(y)-u(x)\big)\,\mathscr{S}(dxdy)\\ &\le \frac{1}{2}\int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\,\alpha^2(x,y)\,\mathscr{S}(dxdy) \;+\; \frac{1}{2}\int_{\overline{\Omega}\times\overline{\Omega}} \frac{\ell^2_{\mathbf{v}}(x,y)}{|y-x|}\,\mathscr{S}(dxdy)\\ &\le \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\left(1+\frac{\alpha^2(x,y)}{2}\right)\mathscr{S}(dxdy) = \mathcal{J}(\pi,\mathscr{S}).\end{aligned}$$

where the first equality is due to property (i) in (5.89) for $\pi = \alpha\mathscr{S}$. It follows that the global extremality condition $\int u\,df = \mathcal{J}(\pi,\mathscr{S})$ holds true if and only the inequalities above are equalities. In view of (5.148) and of Lemma 5.94 this is equivalent to the triple of conditions $(iii), (iv), (v)$. In virtue of Theorem 5.50, equality $\int u\,df = \mathcal{J}(\pi,\mathscr{S})$ characterizes optimal pairs $(\pi,\mathscr{S})$ and $(u,\mathbf{v})$ among the admissible ones, which proves the main part of the theorem.

To prove the "moreover part" we start by recalling the inequality $c_{\mathbf{v}} \le \ell_{\mathbf{v}}$ holding everywhere, see claim (i) of Theorem 5.82. From (5.128) we see that condition (ii) in (5.147) implies that $u(y) - u(x) \le c_{\mathbf{v}}(x,y) \le \ell_{\mathbf{v}}(x,y)$ for $(x,y) \in \overline{\Omega}\times\overline{\Omega}$. The claim (1) readily follows from condition (iii) and the fact that $\ell_{\mathbf{v}}$ and $c_{\mathbf{v}}$ are symmetric. Statements (2), (3) are due to Proposition 5.37 rewritten for $\mathbf{w} = \mathrm{id} - \mathbf{v}$. Indeed, from condition (v) in (5.147) we know that $\langle y'-x', y-x\rangle$ is invariant of the choice of $x' \in \mathbf{v}(x)$, $y' \in \mathbf{v}(y)$ and it equals $\frac{1}{2}\,\ell^2_{\mathbf{v}}(x,y)$ therefore the prerequisite of Proposition 5.37 may be recovered from condition (iii) in (5.147). □

The connection between the saddle point property of $(\overline{\gamma}, \overline{\mathbf{v}})$ and optimality of the string system $(\overline{\pi}, \overline{\mathscr{S}})$ in $(\widetilde{\mathscr{P}^*})$ may be readily stated. The interpretation of this result will be put forth in the forthcoming subsection.

Corollary 5.97. *Let $(\overline{\pi}, \overline{\mathscr{S}})$ and $(\overline{u}, \overline{\mathbf{v}})$ be optimal pairs for, respectively $(\widetilde{\mathscr{P}^*})$ and $(\mathscr{P}_{\mathrm{geo}})$. According to Remark 5.49 we may assume in addition that $\overline{\pi} \le 0$.*

Then the pair $\big(\overline{\mathbf{v}}, \overline{\gamma}\big)$ with $\overline{\gamma} = -\overline{\pi}$ solves the saddle point problem (5.139) for ν_1 and ν_2 being the traces on Ω of the left and right marginals of $\overline{\gamma}$, respectively.

Proof. We shall exploit Theorem 5.95: in particular, we will use the fact that $c_{\overline{\mathbf{v}}}(x,y) = \ell_{\overline{\mathbf{v}}}(x,y) = |\overline{\alpha}(x,y)|\,|y-x|$ for $\mathscr{S}$-a.e. $(x,y) \in \overline{\Omega} \times \overline{\Omega}$ while $\overline{\gamma} = -\overline{\pi} = |\overline{\pi}| = |\overline{\alpha}|\,\overline{\mathscr{S}}$. Thus, taking Proposition 5.55 into account, we obtain

$$\int_{\overline{\Omega}\times\overline{\Omega}} c_{\overline{\mathbf{v}}}\, d\overline{\gamma} = \int_{\overline{\Omega}\times\overline{\Omega}} \ell_{\overline{\mathbf{v}}}\, |\overline{\alpha}|\, d\overline{\mathscr{S}} = \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\, \overline{\alpha}^2(x,y)\, \overline{\mathscr{S}}(dxdy) = \mathcal{Z}. \tag{5.149}$$

On the other hand, by exploiting (5.144) and Young's inequality, for every $\mathbf{v} \in \mathbf{V}_\Omega$ we have:

$$\begin{aligned}\int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}}\, d\overline{\gamma} &\le \int_{\overline{\Omega}\times\overline{\Omega}} \ell_{\mathbf{v}}\, |\overline{\alpha}|\, d\overline{\mathscr{S}} \le \int_{\overline{\Omega}\times\overline{\Omega}} \frac{\ell^2_{\mathbf{v}}(x,y)}{2|y-x|}\, \overline{\mathscr{S}}(dxdy) + \int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\, \frac{\overline{\alpha}^2(x,y)}{2}\, \overline{\mathscr{S}}(dxdy) \\ &\le \int_{\overline{\Omega}\times\overline{\Omega}} |y-x| \left(1 + \frac{\overline{\alpha}^2(x,y)}{2}\right) \overline{\mathscr{S}}(dxdy) = \mathcal{J}\big(\overline{\alpha}\,\overline{\mathscr{S}}, \overline{\mathscr{S}}\big) = \mathcal{Z}. \end{aligned} \tag{5.150}$$

Since $(\overline{\pi}, \overline{\mathscr{S}}) \in \Xi(\Omega, f)$, from condition (i) in (5.89) we may infer that $\nu_1 - \nu_2 = f$. Then, by employing the principle of maximal Monge distance (5.136), we infer that for every $\gamma \in \Gamma^{\partial\Omega}(\nu_1, \nu_2)$:

$$\int_{\overline{\Omega}\times\overline{\Omega}} c_{\overline{\mathbf{v}}}\, d\gamma \;\ge\; W^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(\nu_1, \nu_2) = \mathcal{Z}. \tag{5.151}$$

In view of (5.149), (5.150) and (5.151), we deduce that $(\overline{\gamma}, \overline{\mathbf{v}})$ is a saddle point. □

5.6.5. Geodesics with respect to the $c_{\mathbf{v}}$-metric – discussion

For every maximal monotone map $\mathbf{v} \in \mathbf{V}_\Omega$ the cost function $c_{\mathbf{v}}$ defined in (5.127) is a non-negative, symmetric and sub-additive function on $\overline{\Omega} \times \overline{\Omega}$. It may happen, however that $c_{\mathbf{v}}(x,y) = 0$ for distinct $x \neq y$. The function $c_{\mathbf{v}}$ is thus a pseudo-metric and $(\overline{\Omega}, c_{\mathbf{v}})$ is a pseudo-metric space. A metric quotient space $(X_{\mathbf{v}}, d_{\mathbf{v}})$ may be constructed in a standard way:

$$x \sim x' \quad \Leftrightarrow \quad c_{\mathbf{v}}(x,x') = 0, \qquad X_{\mathbf{v}} := \overline{\Omega}\,/\!\sim, \qquad d_{\mathbf{v}}\big([x],[y]\big) := c_{\mathbf{v}}(x,y),$$

where $[x] = \big\{x' \in \overline{\Omega} : x \sim x'\big\}$ for $x \in \overline{\Omega}$ is the class of equivalence relation $\sim$. It often happens that this relation is non-trivial (i.e. there are $x \in \overline{\Omega}$ for which $[x]$ is not a singleton) when $\mathbf{v}$ maximizes the Monge-Kantorovich cost $W^{\partial\Omega}_{c_{\mathbf{v}}}(\nu_1, \nu_2)$ in (5.136) or, equivalently, together with u it solves $(\mathscr{P}_{\mathrm{geo}})$ or $(u, w) = \big(u, \mathrm{id} - \mathbf{i}(\mathbf{v})\big)$ solves $(\overline{\mathscr{P}})$. We may characterize the classes $[\,\cdot\,]$ for the solutions – analytical or numerical – obtained so far:

Example 5.98. Let us define the univalued map $\mathbf{v} = \mathrm{id} - \mathbf{i}^{-1}(w)$ where, for a disk domain, $w = w_r(r)\,\mathrm{e}_r$ in accordance with (5.78) solves the problem $(\overline{\mathscr{P}})$ in Example 5.46. We can check that $\mathbf{v} = \{0\}$ in the inner disk $D := \big\{|x| \le \frac{R}{2}\big\}$ where u admits a plateau –

consequently $[x] = D$ for any $x \in D$, while one can check that $[x] = \{x\}$ if x belongs to the annulus $\overline{\Omega} \setminus D$. A similar conclusion may be drawn in the four-point problem in Example 5.70 based on the numerical simulation which suggests that $w = \mathrm{id}$ in the inner square Q. A bit less obvious scenario is numerically predicted in Example 5.74: in the inner square $Q = \mathrm{co}\big(\{x_1, x_2, x_3, x_4\}\big)$ only the component $\langle \mathbf{v}, \tau^{x_2,x_4}\rangle$ of the univalued map $\mathbf{v} = \mathrm{id} - \mathbf{i}^{-1}(w)$ appears to be constant on the lines that are parallel to $[x_2, x_4]$. As a consequence $c_{\mathbf{v}}(x, y) \le \ell_{\mathbf{v}}(x, y) = 0$ whenever $[x, y] \subset Q$ and $[x, y] \parallel [x_2, x_4]$. Eventually we may deduce that $[x] = Q \cap \big\{x + t\,\tau^{x_2,x_4} : t \in \mathbb{R}\big\}$ for any $x \in Q$. △

For $(X_{\mathbf{v}}, d_{\mathbf{v}})$ being a metric space by following [7, Chapter 4] we may speak of $d_{\mathbf{v}}$-lengths of curves and address a problem of finding a $d_{\mathbf{v}}$-geodesic connecting a given pair of points (classes) $\big([a], [b]\big) \in X_{\mathbf{v}} \times X_{\mathbf{v}}$. A detailed analysis of such geodesic problem was carried out in the paper [29] co-written by the present author. Here we shall quote some of the results without giving the proofs. This is justified as the purpose of this subsection is mainly to synthesise results that were obtained throughout the whole Section 5.6 and provide their interpretation.

Let us consider a curve in $X_{\mathbf{v}}$ that is parameterized by a function $\Theta : [0, 1] \to X_{\mathbf{v}}$ that is continuous, i.e. with respect to $d_{\mathbf{v}}$ metric. Therefore, if we choose a selection $\theta : [0, 1] \to \overline{\Omega}$ of Θ, namely a function satisfying $\theta(t) \in \Theta(t)$ for each t, it may be discontinuous with respect to Euclidean distance.

Example 5.99. Let us again consider Example 5.46 and the univalued map $\mathbf{v} = \mathrm{id} - \mathbf{i}^{-1}(w)$ where $w = w_r(r)\,\mathrm{e}_r$ is given via (5.78). Let us define the inner closed disk $D := \big\{|x| \le \frac{R}{2}\big\}$ and the closed annulus $A := \big\{\frac{R}{2} \le |x| \le R\big\}$. Whenever $\theta_1 : [0, \frac{1}{2}] \to A$ and $\theta_2 : [\frac{1}{2}, 1] \to A$ are two curves that are continuous with respect to Euclidean metric and moreover $\theta_1(\frac{1}{2}) \in \partial D$ and $\theta_2(\frac{1}{2}) \in \partial D$ then the curve $\Theta : [0, 1] \to X_{\mathbf{v}}$ defined as below

$$\Theta(t) = \begin{cases} \{\theta_1(t)\} & \text{if } t < \frac{1}{2}, \\ D & \text{if } t = \frac{1}{2}, \\ \{\theta_2(t)\} & \text{if } t > \frac{1}{2} \end{cases}$$

is continuous with respect to $d_{\mathbf{v}}$ metric. Meanwhile, any selection $\theta : [0, 1] \to \overline{\Omega}$ of Θ is discontinuous at $t = \frac{1}{2}$ where $\theta(\frac{1}{2}) \in D$. △

The length of a curve Θ may be defined in two equivalent ways:

$$\begin{aligned} L_{\mathbf{v}}(\Theta) &= \sup\left\{\sum_{i=0}^{n-1} d_{\mathbf{v}}\big(\Theta(t_{i+1}), \Theta(t_i)\big) \;:\; 0 \le t_0 < t_1 < \cdots < t_n \le 1\right\} \qquad (5.152) \\ &= \sup\left\{\sum_{i=0}^{n-1} c_{\mathbf{v}}(x_{i+1}, x_i) \;:\; 0 \le t_0 < t_1 < \cdots < t_n \le 1, \;\; x_i \in \Theta(t_i)\right\}, \end{aligned}$$

see [7, Chapter 4]; note that in the second definition the sum does not depend on the choice of $x_i \in \Theta(t_i)$. The problem of finding a geodesic curve in the metric space can be

readily posed: for $(a,b) \in \overline{\Omega}$ we define

$$d^g_{\mathbf{v}}\big([a],[b]\big) := \inf\Big\{ L_{\mathbf{v}}(\Theta) \,:\, \Theta \text{ is a continuous curve in } X_{\mathbf{v}},\ \Theta(0)=[a],\ \Theta(1)=[b]\Big\}, \tag{5.153}$$

where $d^g_{\mathbf{v}}\big([a],[b]\big)$ is called a *geodesic distance* or *inner distance*. Let us stress that *a priori* only inequality $d^g_{\mathbf{v}} \geq d_{\mathbf{v}}$ is assured to hold true. If the equality holds for any pair $[a],[b]$ then we say that $(X_{\mathbf{v}}, d_{\mathbf{v}})$ is a *geodesic metric space*.

For a given $\mathbf{v} \in \mathbf{V}_\Omega$ we may in addition define a *Finsler-type* local metric $\varphi_{\mathbf{v}} : \overline{\Omega}\times\mathbb{R}^d \to \overline{\mathbb{R}}$ given as

$$\varphi_{\mathbf{v}}(x,z) = \limsup_{h\to 0_+} \frac{1}{h}\, c_{\mathbf{v}}\big(x, x+hz\big). \tag{5.154}$$

It is a Borel function which is convex, lower semi-continuous and one-homogeneous in z, see [165].

Proposition 5.100. *Let* $\mathbf{v}$ *be a monotone map in the set* $\mathbf{V}_\Omega$*. Then:*

(i) The quotient space $(X_{\mathbf{v}}, d_{\mathbf{v}})$ *is a geodesic locally compact metric space, in particular* $d^g_{\mathbf{v}}\big([a],[b]\big) = d_{\mathbf{v}}\big([a],[b]\big)$ *for any* $(a,b) \in \overline{\Omega}\times\overline{\Omega}$*.*

(ii) For any pair $(a,b) \in \overline{\Omega}\times\overline{\Omega}$ *the geodesic problem attains a solution* $\hat{\Theta} : [0,1] \to X_{\mathbf{v}}$*. Moreover* $\hat{\Theta}$ *can be chosen so that* $d_{\mathbf{v}}\big(\hat{\Theta}(s), \hat{\Theta}(t)\big) = d_{\mathbf{v}}([a,b])\,|t-s|$ *for any* $0 \leq s \leq t \leq 1$*.*

(iii) For a given pair $(a,b) \in \overline{\Omega}\times\overline{\Omega}$ *assume that the infimum in* (5.127) *is attained for suitable points* $x_1, x_2, \ldots, x_N \in \overline{\Omega}$*. Then the polygonal curve* $\mathrm{C} = \cup_{i=1}^{N-1}[x_i, x_{i+1}]$ *is a geodesic joining* a *to* b *while for each* i *the scalar function* $\mathbf{v}_{x_i,x_{i+1}}$ *is affine on* $[0,1]$*.*

(iv) Assume that $\mathbf{v}(x) = \{v(x)\}$ *where* v *is Lipschitz continuous. Then, for any curve* Θ *such that* $\Theta(t) = \{\theta(t)\}$ *for* $\theta \in \mathrm{Lip}([0,1];\mathbb{R}^d)$ *we have*

$$L_{\mathbf{v}}(\Theta) = \int_0^1 \varphi_{\mathbf{v}}\big(\theta(t), \theta'(t)\big)\, dt. \tag{5.155}$$

In addition then we have

$$\varphi_{\mathbf{v}}(x,z) = \sqrt{2}\,\Big(\langle e(v)(x), z\otimes z\rangle\Big)^{1/2}$$

at every point x *of differentiability of* v*.*

Proof. See [29, Proposition 5.6 and Corollary 5.9]. □

In what follows we shall analyse the functions $\mathbf{v}$ that were optimal in examples studied in this chapter. We will explain why those functions furnish "the worst" Monge-Kantorovich cost $c_{\mathbf{v}}$ in the sense of the problem (5.136). We will examine the optimal transportation plans γ with respect to that cost, i.e. that solve $(\mathrm{OTP}^{\partial\Omega}_{c_{\mathbf{v}}})$. For non-negative load f this will be connected to identifying the shortest geodesics connecting the points in the support of f to boundary points. Finally, we will characterize the subset of optimal plans γ that incorporates the condition of $(\gamma, \mathbf{v})$ being a saddle point for the Lagrangian (5.138). Often, provided it does not cause confusion, we will abuse the language and speak of $c_{\mathbf{v}}$ metric instead of $d_{\mathbf{v}}$ metric and about geodesics on $\overline{\Omega}$ instead on $X_{\mathbf{v}}$, etc.

Example 5.101 (**The one-point-force load revisited**)**.** Let us begin with the one point load $f = \delta_{x_0}$ and a disk domain $\Omega = \big\{|x| < R\big\}$ for which in Section 5.5.3 we have derived analytical solutions $(\overline{u}, \overline{w})$ and $(\overline{\pi}, \overline{\mathscr{S}})$ of problems $(\overline{\mathscr{P}})$ and $(\widetilde{\mathscr{P}^*})$ respectively, therefore the univalued map $\overline{\mathbf{v}}(x) = \{\overline{v}(x)\} = \{x - \overline{w}(x)\}$ together with $\overline{u}$ solves $(\mathscr{P}_{\mathrm{geo}})$ by virtue of Proposition 5.84. Let us recall that $\overline{v}$ is a Lipschitz continuous function that on each segment $[x_0, a]$ with $a \in \partial\Omega$ is affine and changes from $\overline{v}(x_0) = x_0 - \overline{w}(x_0) = -x_0$ to $\overline{v}(a) = a$. Moreover, we had $\overline{\pi} = -\delta_{x_0} \otimes \overline{p}$ (we may change the sign in accordance with Remark 5.49) and $\overline{\mathscr{S}} = \delta_{x_0} \otimes \overline{s}$ where $\overline{p}$ is a probability on $\partial\Omega$ that satisfies $x_0 = \int_{\partial\Omega} y\, \overline{p}(dy)$ while $\overline{s} \in \mathcal{M}_+(\partial\Omega)$ and $\overline{s}(dy) = \frac{|x_0 - y|}{\sqrt{2}\, d_0}\, \overline{p}(dy)$ where $d_0 = \sqrt{R^2 - |x_0|^2}$. Fig. 5.4(a) conveys the freedom of choice of $\overline{p}$.

We may choose an optimal $\overline{p}$ so that $\operatorname{spt} \overline{p} = \partial\Omega$ resulting in $\operatorname{spt}(\overline{\pi}, \overline{\mathscr{S}}) = \bigcup_{a \in \partial\Omega} (x_0, a)$. Then, by the "moreover part" of Theorem 5.95 we know that $c_{\overline{\mathbf{v}}}(x_0, a) = \ell_{\overline{\mathbf{v}}}(x_0, a)$ for every $a \in \partial\Omega$. Accordingly

$$\begin{aligned} c_{\overline{\mathbf{v}}}(x_0, a) = \ell_{\overline{\mathbf{v}}}(x_0, a) &= \sqrt{2\,\langle \overline{v}(x_0) - \overline{v}(a), x_0 - a\rangle} = \sqrt{2\,\langle -x_0 - a, x_0 - a\rangle} \\ &= \sqrt{2}\sqrt{|a|^2 - |x_0|^2} = \sqrt{2}\sqrt{R^2 - |x_0|^2} = \sqrt{2}\, d_0 = \mathcal{Z}, \end{aligned} \tag{5.156}$$

i.e. the optimality of function $\overline{\mathbf{v}}$ guarantees that all boundary points are of equal $c_{\overline{\mathbf{v}}}$-distance to the point x_0. On top of that, owing to assertion (iii) in Proposition 5.100, the equality $c_{\overline{\mathbf{v}}}(x_0, a) = \ell_{\overline{\mathbf{v}}}(x_0, a)$ guarantees (take $N = 2$ in the assertion) that for every $a \in \partial\Omega$ the straight segments $[x_0, a]$ are $c_{\overline{\mathbf{v}}}$-geodesics that connect pairs of points (x_0, a) and the geodesics are all of equal length $c_{\overline{\mathbf{v}}}(x_0, a) = \sqrt{2}\, d_0$.

Next we make use of Corollary 5.97. We see that $\overline{\pi} \le 0$ and that restrictions to Ω of the left and right marginals of $\overline{\gamma} := -\overline{\pi} \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ are equal to, respectively, $\nu_1 = \delta_{x_0} = f$ and $\nu_2 = p \llcorner \Omega = 0$. Thus from Corollary 5.97 we deduce that the pair $(\overline{\gamma}, \overline{\mathbf{v}})$ is a saddle point for the Lagrangian (5.138). In particular $\overline{\gamma}$ solves problem $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$. Independently, this is clear by Proposition 5.87 since $\{x_0\} \times \partial\Omega$ is a subset of $\mathcal{G}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}$ owing to the fact that $c_{\overline{\mathbf{v}}}$ is constant on this set. In turn, the same statement assures that $\gamma = \delta_{x_0} \otimes p$ is optimal in $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ for any probability $p \in \mathrm{P}(\partial\Omega)$, i.e. it is not obligatory to meet the barycentre condition $x_0 = \int_{\partial\Omega} y\, p(dy)$. As it will turn out, the latter condition cannot be skipped if we wish that $(\gamma, \overline{\mathbf{v}})$ is a saddle point for the Lagrangian (5.138). Indeed, let $a_1, a_2 \in \partial\Omega$ be the boundary points such that $[a_1, a_2]$ is the diameter of Ω for which $x_0 \in [a_1, a_2]$ and $|x_0 - a_1| < |x_0 - a_2|$ (we assume that x_0 is not the origin). The "one-ray" transport plan $\tilde{\gamma} = \delta_{x_0} \otimes \delta_{a_1}$ solves $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ but $\sup_{\mathbf{v} \in \mathbf{V}_\Omega} \int c_{\mathbf{v}} d\tilde{\gamma}$ is larger than $\int c_{\overline{\mathbf{v}}} d\tilde{\gamma} = \int c_{\overline{\mathbf{v}}} d\overline{\gamma} = \mathcal{Z}$. We can propose an alternative univalued map $\tilde{\mathbf{v}} = \{\tilde{v}\}$ and we may concentrate on maximizing $c_{\tilde{\mathbf{v}}}$ on the support of $\tilde{\gamma}$ only, i.e. on the pair (x_0, a_1). While keeping the affine character of the monotone map on the rays $[x_0, a]$ (we still have $c_{\tilde{\mathbf{v}}}(x_0, a) = \ell_{\tilde{\mathbf{v}}}(x_0, a)$, although we skip the proof here) we put $\tilde{v}(x_0) = a_2$ and this way we maximize the tangential slope of $\tilde{\mathbf{v}}$ (among possible $\tilde{\mathbf{v}} \in \mathbf{V}_\Omega$) on $[x_0, a_1]$ without minding the slope on $[x_0, a_2]$ where it is now actually zero. Consequently $c_{\tilde{\mathbf{v}}}(x_0, a_1) = \sqrt{2\,\langle a_2 - a_1, x_0 - a_1\rangle} > \mathcal{Z}$ while $c_{\tilde{\mathbf{v}}}(x_0, a_2) = 0$ and thus $\int c_{\tilde{\mathbf{v}}} d\tilde{\gamma} > \mathcal{Z}$.

Let us now write down some implications of $(\overline{\gamma}, \overline{\mathbf{v}})$ being a saddle point for (5.138). Since we are interested only in only *some* of the implications we may consider variations

amongst $\gamma = \delta_{x_0} \otimes p$ with $p \in \mathrm{P}(\partial\Omega)$ and univalued $\mathbf{v} = \{v\} \in \mathbf{V}_\Omega$ changing affinely on the rays $[x_0, a]$. Such maps $\mathbf{v}$ are completely characterized by $v(x_0) =: v_0 \in \mathbb{R}^2$, namely $\mathbf{v} = \mathbf{v}(v_0)$. Eventually we are considering a simpler Lagrangian:

$$\mathscr{L}(p, v_0) := \int_{\overline{\Omega}\times\overline{\Omega}} c_{\mathbf{v}(v_0)}(x, y)\, \gamma(dxdy) = \int_{\partial\Omega} c_{\mathbf{v}(v_0)}(x_0, y)\, p(dy) = \int_{\partial\Omega} \ell_{\mathbf{v}(v_0)}(x_0, y)\, p(dy).$$

The necessary conditions for $\overline{p}, \overline{v}_0$ to be a stationary points are the following:

(i) the stationary condition with respect to p: there exists $\lambda \in \mathbb{R}_+$ (the Lagrange multiplier for the constraint $p \in \mathrm{P}(\partial\Omega)$) such that

$$\begin{cases} \ell_{\mathbf{v}(\overline{v}_0)}(x_0, y) \geq \lambda & \text{for every } y \in \partial\Omega, \\ \ell_{\mathbf{v}(\overline{v}_0)}(x_0, y) = \lambda & \text{for } \overline{p}\text{-a.e. } y \in \partial\Omega; \end{cases} \tag{5.157}$$

(ii) the stationary condition with respect to v_0:

$$\int_{\partial\Omega} \frac{x_0 - y}{\ell_{\mathbf{v}(\overline{v}_0)}(x_0, y)}\, \overline{p}(dy) = 0 \tag{5.158}$$

(recall that $\ell_{\mathbf{v}(v_0)}(x_0, y) = \sqrt{2\,\langle v_0 - y, x_0 - y\rangle}$).

From (5.157) we can see that in the denominator in (5.158) we have a constant. Readily, the barycentre condition for $\overline{p}$ is recovered as a necessary condition for $\overline{\gamma} = \delta_{x_0} \otimes \overline{p}$ to be the first component of a saddle point for (5.138).

To the stationary condition (5.158) we may assign an interpretation that is directly linked to the problem of finding an optimal string system. We know that $\gamma = \delta_{x_0} \otimes p$ is optimal in $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ for any $p \in \mathrm{P}(\partial\Omega)$. Let us define $\pi = -\gamma$ and then ask: what additional condition p must satisfy so that there exists $\mathscr{S} \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ such that $(\pi, \mathscr{S})$ solves $(\widetilde{\mathscr{P}}^*)$? Secondly, can such $\mathscr{S}$ be computed directly from π or γ? To give the answer we immediately turn to the optimality conditions in (5.147). Clearly $\pi = \alpha\mathscr{S}$ therefore $\mathscr{S} = \delta_{x_0} \otimes s$ with $p(dy) = \alpha(x_0, y)\, s(dy)$. From conditions (iii), (iv) we find that $\alpha(x_0, y) = -\frac{\ell_{\overline{\mathbf{v}}}(x_0, y)}{|y - x_0|}$ for s-a.e. y. Since $\ell_{\overline{\mathbf{v}}}(x_0, y) > 0$ for every $y \in \partial\Omega$ we may compute $s(dy) = \frac{|y-x_0|}{\ell_{\overline{\mathbf{v}}}(x_0, y)}\, p(dy)$ or

$$\mathscr{S}(dxdy) = -\frac{|y - x|}{\ell_{\overline{\mathbf{v}}}(x, y)}\, \pi(dxdy) = \frac{|y - x|}{\ell_{\overline{\mathbf{v}}}(x, y)}\, \gamma(dxdy). \tag{5.159}$$

The pre-stress $\mathscr{S}$ is thus already determined for a fixed γ, while all optimality conditions in (5.147) are satisfied except condition (i), more accurately it is left to ensure that $\int_{\overline{\Omega}\times\overline{\Omega}} \langle \phi(y) - \phi(x), \tau^{x,y} \rangle\, \mathscr{S}(dxdy) = 0$ for any $\phi \in C_0(\Omega; \mathbb{R}^2)$. For $\mathscr{S}$ of the particular form herein considered the condition boils down to $\int_{\partial\Omega} \frac{y - x_0}{|y - x_0|}\, s(dy) = 0 \in \mathbb{R}^2$ being the in-plane equilibrium equation of the joint at x_0. After plugging the formula for $s(dy)$ derived above into this equation we recast the stationary condition (5.158). This condition on $p = \overline{p}$ thus guarantees existence of $\mathscr{S}$ such that $(\pi, \mathscr{S}) = (-\gamma, \mathscr{S})$ is a solution of $(\widetilde{\mathscr{P}}^*)$ while (5.159) is the formula for such $\mathscr{S}$.

The study above translates to a large extent to the one-point-force load $f = \delta_{x_0}$ in domains other than disks, see e.g. the case of a rectangular domain Ω in Example 5.62. The difference is that in general not all boundary points $a \in \partial\Omega$ are of the same $c_{\overline{\mathbf{v}}}$-distance

from x_0, i.e. satisfy $c_{\overline{\mathbf{v}}}(x_0,a) = d^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(x_0)$. These are the points at which strings of the optimal systems end: for the rectangular domain we have either two such points – Fig. 5.5(a) or (c), or three such points – Fig. 5.5(b), or finally four such points if Ω is a square – Fig. 5.1(b). Apart from the last scenario, the stationary conditions (5.157) and (5.158) yield a unique $p = \overline{p}$ and therefore unique $\overline{\gamma}$ and $\overline{\mathscr{S}}$, cf. (5.159).

*Summarizing, the following observations can be made for a one-point-force load $f = \delta_{x_0}$ (**and for the one-point-force load only**):*

1. *The maximizer $\overline{\mathbf{v}} \in \mathbf{V}_\Omega$ for the problem (5.136) changes the Euclidean metric to $c_{\overline{\mathbf{v}}}$-metric so that effectively the smallest distance $c_{\overline{\mathbf{v}}}(x_0,a) = d^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(x_0)$ is realized for more than one boundary point $a \in \partial\Omega$. In the special case of a disk domain Ω this set of points – namely $\Pi^{\partial\Omega}_{c_{\mathbf{v}}}(x_0)$ – is equal to the whole boundary $\partial\Omega$.*
2. *Owing to the aforementioned property of $\overline{\mathbf{v}}$ there are always multiple transport plans γ that are optimal in $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$. They are of the form $\gamma = \delta_{x_0} \otimes p$ where $p \in \mathrm{P}\big(\Pi^{\partial\Omega}_{c_{\mathbf{v}}}(x_0)\big)$.*
3. *Among all the optimal γ constructed as above there is at least one plan $\overline{\gamma}$ such that $(\overline{\gamma}, \overline{\mathbf{v}})$ is a saddle point for the Lagrangian (5.138). For each such optimal plan $\overline{\gamma}$ a string system $(\overline{\pi}, \overline{\mathscr{S}})$ that is optimal for $(\widetilde{\mathscr{P}^*})$ may be computed as below:*

$$\overline{\pi}(dxdy) = -\overline{\gamma}(dxdy), \qquad \overline{\mathscr{S}}(dxdy) = \frac{|y-x|}{\ell_{\overline{\mathbf{v}}}(x,y)}\,\overline{\gamma}(dxdy). \tag{5.160}$$

△

The path from the saddle point $(\overline{\mathbf{v}}, \overline{\gamma})$ to an optimal string system $(\overline{\pi}, \overline{\mathscr{S}})$ gets corrupted for the more complicated loads:

Example 5.102 (**Four-point-forces load revisited**)**.** Let us now go back to the four-forces problem for which the solutions $(\overline{u}, \overline{w})$ and $(\overline{\pi}, \overline{\mathscr{S}})$ were numerically predicted in Example 5.70. Let us agree to speak of this solution as if it was exact. We put $\overline{\mathbf{v}}(x) = \{x - \overline{w}(x)\}$ as the univalued maximal monotone map in $\mathbf{V}_\Omega$. Recall that $\overline{\mathbf{v}} = 0$ in the inner square $Q = \mathrm{co}\big(\{x_1, x_2, x_3, x_4\}\big)$ (see Example 5.98) therefore Q is a single point (an equivalence class) in the quotient metric space $X_{\overline{\mathbf{v}}}$.

By arguing similarly as in Example 5.101 we deduce that $c_{\overline{\mathbf{v}}}(x_i, a_{ij}) = \ell_{\overline{\mathbf{v}}}(x_i, a_{ij})$ for each i, j, see Fig. 5.6(a). Accordingly, the segments $[x_i, a_{ij}]$ are $d_{\overline{\mathbf{v}}}$-geodesics or more precisely $\Theta_{ij} : [0,1] \ni t \mapsto [x_i + t(a_{ij} - x_i)] \in X_{\overline{\mathbf{v}}}$ are parametrizations of geodesic curves, while $\Theta_{ij}(0) = Q$. We thus observe that, for instance, selection θ_{11} of Θ_{11} can be continuous, e.g. $\theta_{11} : [0,1] \ni t \mapsto \big(x_1 + t(a_{11} - x_1)\big) \in \overline{\Omega}$, or discontinuous, e.g. $\theta_{11} :]0,1] \ni t \mapsto \big(x_1 + t(a_{11} - x_1)\big) \in \overline{\Omega}$ and $\theta_{11}(0) = x_3$. As a result we obtain that the distances $c_{\overline{\mathbf{v}}}(x_i, a_{11})$ are equal for each i or similarly $c_{\overline{\mathbf{v}}}(x_1, a_{ij})$ are equal for all i, j.

By Corollary 5.97 the transport plan $\overline{\gamma} = -\overline{\pi}$ (with $\overline{\pi} = \pi_h$, see Example 5.70) together with the map $\overline{\mathbf{v}}$ constitute a saddle point for the Lagrangian (5.138) and in particular $\overline{\gamma}$ solves the $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ problem for the marginals $\nu_1 = f$, $\nu_2 = 0$. Considering the analysis above we deduce that $\Pi^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(x) = \{a_{ij}\}_{i=1,j=2}^{i=4,j=2}$ for every $x \in Q$, therefore the set of optimal plans $\mathcal{M}_+\big(\mathcal{G}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}\big)$ offers a lot of possibilities. For example, according to Proposition 5.87,

the plan $\gamma = \sum_{i=1}^{4} \delta_{(x_i,a_{11})}$ is also optimal for $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$, namely all the mass at points $x_1, \ldots, x_4$ may be transported to a selected boundary point, a_{11} for instance.

Despite the non-uniqueness of solution γ to the problem $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ one may show that $\overline{\gamma} = -\overline{\pi}$ is the only transport plan such that $(\overline{\gamma}, \overline{\mathbf{v}})$ is a saddle point – this plan uniformly distributes mass at each x_i to boundary points a_{i1} and a_{i2}, see Fig. 5.6(b) for the graphical representation. Can we now follow the footsteps of remark (3) in Example 5.102 and recast optimal $\overline{\mathscr{S}}$ from $\overline{\gamma}$ using the second formula in (5.160)? Sadly, we have $\ell_{\overline{\mathbf{v}}}(x, y) = 0$ for all pairs $(x, y) \in Q \times Q$ therefore the latter formula cannot be employed there. Nevertheless, the saddle point property of $\overline{\gamma}$ assures that there exists a pre-stress measure $\overline{\mathscr{S}}$ such that $(-\overline{\gamma}, \overline{\mathscr{S}})$ is an element of $\Xi(\Omega, f)$ and moreover $\overline{\gamma}(dxdy) = \frac{\ell_{\overline{\mathbf{v}}}(x,y)}{|y-x|}\,\overline{\mathscr{S}}(dxdy)$. We can therefore compute $\overline{\mathscr{S}}$ on the set $\bigcup_{i=1,j=1}^{i=4,j=2} \{(x_i, a_{ij})\}$ where $\ell_{\overline{\mathbf{v}}} > 0$. Contrarily, on the set $Q \times Q$ there is no local condition on $\mathscr{S}$ – our task is to construct it so that the in-plane equilibrium (ii) in (5.89) is met. There are plenty of ways to achieve that, which is reflected in the cloud of strings in Fig. 5.6(a). △

Example 5.103 (**Five-point-forces load revisited**)**.** This time we shall analyse the metric induced by the univalued monotone map $\overline{\mathbf{v}}(x) = \{x - \overline{w}(x)\}$ in the case of the five-point-forces load numerically solved in Example 5.71. Once again let us speak about the numerical predictions (u_h, w_h) and $(\pi_h, \mathscr{S}_h)$ as if they were exact solutions $(\overline{u}, \overline{w})$ and $(\overline{\pi}, \overline{\mathscr{S}})$. Again we may utilize Corollary 5.97 to deduce that $(\overline{\gamma}, \overline{\mathbf{v}})$ is a saddle point for Lagrangian (5.138) where

$$\overline{\gamma} = -\overline{\pi} = \sum_{i=1}^{4} \frac{1}{4}\,\delta_{(x_0,x_i)} + \sum_{i=1}^{4}\sum_{j=1}^{2} \frac{5}{8}\,\delta_{(x_i,a_{ij})},$$

cf. Fig. 5.7(b) for the graphical display of this transport plan. In particular $\overline{\gamma}$ solves $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ but this time the source and target measures (which are traces of marginals of $\overline{\gamma}$ on Ω) are: $\nu_1 = \delta_{x_0} + \sum_{i=1}^{4} \frac{5}{4}\,\delta_{x_i}$ and, respectively, $\nu_2 = \sum_{i=1}^{4} \frac{1}{4}\,\delta_{x_i}$. This is a new scenario comparing to the one-force load and the four-force load, i.e. ν_1 is not equal to f and ν_2 is non-zero, although $\nu_1 - \nu_2 = f$. Based on optimality conditions (5.147) once again we may deduce that $c_{\overline{\mathbf{v}}}(x_0, x_i) = \ell_{\overline{\mathbf{v}}}(x_0, x_i) > 0$ for each $i \geq 1$ and $c_{\overline{\mathbf{v}}}(x_i, a_{ij}) = \ell_{\overline{\mathbf{v}}}(x_i, a_{ij}) > 0$ for each $i \geq 1$ and j. Accordingly, the corresponding segments $[x_0, x_i]$ and $[x_i, a_{ij}]$ are $c_{\overline{\mathbf{v}}}$-geodesics. Similarly as in the case of one-force load we can retrieve $\overline{\mathscr{S}}$ from $\overline{\gamma}$ by exploiting (5.160).

Notwithstanding the above we may always readdress the problem $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ while enforcing that $\widetilde{\nu}_1 = f$ and $\widetilde{\nu}_2 = 0$ after all – since $\widetilde{\nu}_1 - \widetilde{\nu}_2 = f$ the Monge-Kanotorvich cost $W^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(\widetilde{\nu}_1, \widetilde{\nu}_2)$ is still equal to $\mathcal{Z}$. We propose the transport plan whose marginals are precisely $\widetilde{\nu}_1$ and $\widetilde{\nu}_2$:

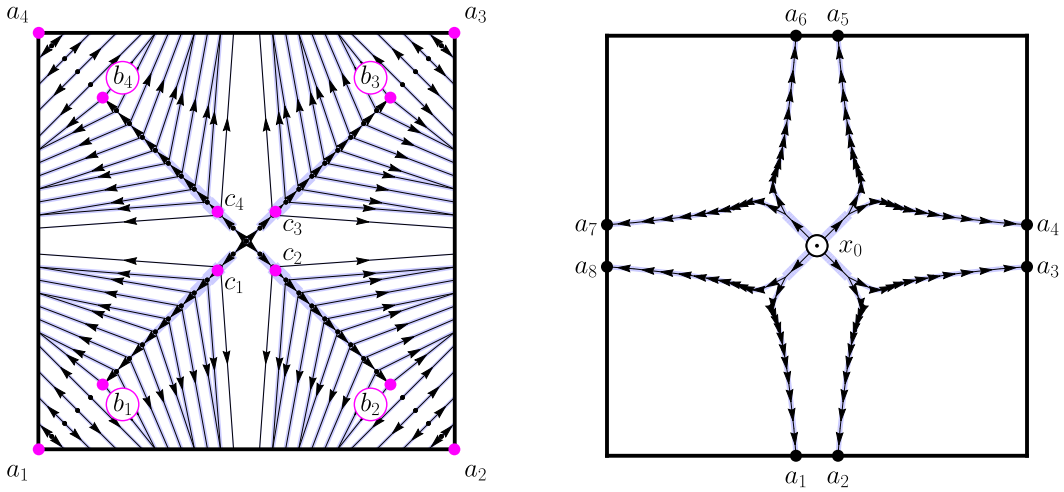
$$\widetilde{\gamma} = \sum_{i=1}^{4}\sum_{j=1}^{2} \frac{1}{8}\,\delta_{(x_0,a_{ij})} + \sum_{i=1}^{4}\sum_{j=1}^{2} \frac{1}{2}\,\delta_{(x_i,a_{ij})},$$

i.e. the mass x_0 is equally partitioned and directly transported to the eight points a_{ij}. To check if $\widetilde{\gamma}$ is optimal for $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ we must verify whether $\int c_{\overline{\mathbf{v}}}\, d\widetilde{\gamma} = \mathcal{Z}$. To that aim we must compute $c_{\overline{\mathbf{v}}}(x_0, a_{ij})$. By the very definition (5.127) we have $c_{\overline{\mathbf{v}}}(x_0, a_{ij}) \leq \ell_{\overline{\mathbf{v}}}(x_0, x_i) +$

$\ell_{\overline{\mathbf{v}}}(x_i, a_{ij}) = c_{\overline{\mathbf{v}}}(x_0, x_i) + c_{\overline{\mathbf{v}}}(x_i, a_{ij})$. Thus it is easy to check that $\int c_{\overline{\mathbf{v}}}\, d\widetilde{\gamma} \le \int c_{\overline{\mathbf{v}}}\, d\overline{\gamma} = \mathcal{Z} = W^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(\widetilde{\nu}_1, \widetilde{\nu}_2)$ therefore $\widetilde{\gamma}$ must be optimal while the aforementioned inequalities must be equalities. By the claim (iii) in Proposition 5.100 we infer that the polygonal chains $[x_0, x_i] \cup [x_i, a_{ij}]$ are the eight $c_{\overline{\mathbf{v}}}$-geodesics connecting x_0 with respective boundary points a_{ij}. These curves have the minimal $c_{\overline{\mathbf{v}}}$-length among all possible curves connecting x_0 to $\partial\Omega$. It is remarkable that the metric $c_{\overline{\mathbf{v}}}$ is chosen so that those minimal curves are piece-wise straight while their vertices are in $\operatorname{spt} f$. △

From the previous examples we may heuristically extract a method of predicting $c_{\overline{\mathbf{v}}}$-geodesics that connect the points x in the support of $f \ge 0$ to points $a \in \partial\Omega$ with the shortest possible length $c_{\overline{\mathbf{v}}}(x, a)$. Namely, it is enough to analyse the picture of ϑ_π (if an optimal string system $(\pi, \mathscr{S})$ exists and is available) or the numerical simulation ϑ_{π_h}, see for instance Fig. 5.7(b). Starting from a particular point $x \in \operatorname{spt} f$ we ought to track the polygonal chains connecting x to a point on $\partial\Omega$ while taking the direction of $-\vartheta_{\pi_h}$ into consideration. We may employ this technique for the next example:

Example 5.104 (**Diagonal load revisited**)**.** We will now study the $c_{\overline{\mathbf{v}}}$-geodesics for the univalued map $\overline{\mathbf{v}}(x) = \{x - \overline{w}(x)\}$ with $\overline{w}$ numerically computed in the optimal membrane problem with the load distributed along the diagonals of a square domain Ω, see Example 5.72. We append the numerical solutions presented in Fig. 5.8 by displaying ϑ_{π_h} in Fig. 5.13(a) (we use a lower resolution h for better readability). Based on the numerical solution



(a) (b)

Figure 5.13: (a) Numerical prediction ϑ_{π_h} for the diagonal load considered in Example 5.72; (b) eight equivalent $c_{\overline{\mathbf{v}}}$-geodesics connecting x_0 to the boundary with the smallest $c_{\overline{\mathbf{v}}}$-length – a numerical prediction for $\overline{\mathbf{v}} = \mathrm{id} - \mathbf{i}^{-1}(\overline{w})$ computed for the uniform pressure load in Example 5.73.

ϑ_{π_h} we can predict the structure of $c_{\overline{\mathbf{v}}}$-geodesics connecting points in $\operatorname{spt} f$ to boundary points. To this aim it is crucial to observe that ϑ_{π_h} charges segments $[b_1, b_3]$, $[b_2, b_4]$. With x_0 denoting the centre of the square we can foresee that:

(i) for $\mathcal{H}^1$-a.e. $x \in [x_0, c_i]$ the geodesics may be characterized as polygonal chains

$$[x, z] \cup [z, \bar{y}_1(z)] \qquad \text{or} \qquad [x, z] \cup [z, \bar{y}_2(z)]$$

where z is an arbitrary element in $[c_i, b_i]$ while $\bar{y}_1(z), \bar{y}_2(z) \in \partial\Omega$ are the boundary points uniquely determined for each such z and positioned symmetrically with respect to diagonal ending at a_i;

(ii) for $\mathcal{H}^1$-a.e. $x \in [c_i, b_i]$ the geodesic may be either a single segment or again a polygonal chain:

$$[x, \bar{y}_1(x)] \quad \text{or} \quad [x, \bar{y}_2(x)] \qquad \text{or} \qquad [x, z] \cup [z, \bar{y}_1(z)] \quad \text{or} \quad [x, z] \cup [z, \bar{y}_2(z)]$$

where z is an arbitrary element in $]x, b_i]$;

(iii) for $\mathcal{H}^1$-a.e. $x \in [b_i, a_i]$ the geodesics are segments

$$[x, \hat{y}_1(x)] \quad \text{or} \quad [x, \hat{y}_2(x)]$$

where $\hat{y}_1(x), \hat{y}_2(x) \in \partial\Omega$ are projections of x onto the boundary along the direction that is orthogonal to the diagonal ending at a_i.

These predictions were confirmed numerically, i.e. by solving a suitable optimal transshipment problem – see Example 5.105 below for details. The (presumed) freedom of choosing the geodesics for points in $[x_0, b_i]$ is remarkable – a variant of observation (1) in Example 5.101 can be made, although now the structure of the geodesic is far less obvious. △

Example 5.105 (**Pressure load revisited**)**.** So far all the $c_{\overline{\mathbf{v}}}$-geodesics (computed analytically or predicted numerically) were either straight segments or polygonal chains. We shall now show that this seems not to be the case for the pressure load $f = \mathcal{L}^2 \llcorner \Omega$ considered in Example 5.73. The method for recognizing the geodesics that was employed in Example 5.104 would be very difficult to apply here: the graphical representation of ϑ_{π_h} tightly fills the whole domain Ω, similarly as $\varsigma_{\mathscr{S}_h}$ does in Fig. 5.9(a).

Instead we may propose a more consistent method of predicting a $c_{\overline{\mathbf{v}}}$-geodesic for a pair of points $(a, b) \in \mathrm{X}^h \times \mathrm{X}^h$. It consists in solving a numerical approximation of the *optimal transshipment problem* for source and target measures $\nu_1 = \delta_a$ and, respectively, $\nu_2 = \delta_b$ with the cost function being $\ell_{\overline{\mathbf{v}}}$ which is straightforward to compute in contrast to $c_{\overline{\mathbf{v}}}$. The difference between the optimal transport problem and the optimal transshipment problem is that in the latter we impose the condition on the difference of marginals: $\pi_1\#\gamma - \pi_2\#\gamma = \nu_1 - \nu_2$ rather than on the two marginals independently, see [166, Section 1.2.2]. Solution of the transshipment problem will not exists in general while the corresponding value function is precisely $c_{\overline{\mathbf{v}}}(a, b)$ (we skip a detailed argumentation). If we additionally enforce that both marginals are supported on X^h we get ourselves a ground-structure-based numerical method for such a transshipment problem: we obtain a numerical estimate $c_{\overline{\mathbf{v}}}(a, b)$ from above, while from the support of optimal γ we find polygonal-chain approximations of the $c_{\overline{\mathbf{v}}}$-geodesics connecting a and b.

The transshipment problem may be easily modified so that there are hidden source/target measure on $\partial\Omega$, cf. the problem $(\mathrm{OTP}^{\partial\Omega}_{c_{\mathbf{v}}})$. This way, if we set $\nu_1 = \delta_{x_0}$ we obtain numerical approximations of geodesics connecting x_0 and $a \in \partial\Omega$ with the smallest

length $c_{\overline{\mathbf{v}}}(x_0, a)$. Let us now take the numerical approximation of $\overline{\mathbf{v}}(x) = \{\mathrm{id} - \overline{w}(x)\}$ computed for the pressure load in Example 5.73 (at the nodes in X^h only) and solve such a modified transshipment problem for x_0 being the centre of the square. In Fig. 5.13(b) we can see the graphical representation of numerical solution γ_h or rather ϑ_{γ_h}. Based on this figure we may foresee that the polygonal chains that connect x_0 to either of the eight points a_i are approximations of geodesics, which clearly are curved. △

5.7. Open problems

The optimal membrane problem is a new topic that was put forth in a joint work by the present author and Guy Bouchitté. Although the core theory of the problem has been established and a lot of important issues has been resolved, there are still plenty of unanswered questions. We wrap up this chapter by collecting the most interesting of those.

5.7.1. Finitely supported load and existence of an optimal string system

Despite the lack of solutions to problem $(\widetilde{\mathscr{P}^*})$ in general, i.e. of optimal string systems (see Remark 5.59), there are strong indications that such a solution exists for a discrete load f, namely when $\operatorname{spt} f$ is finite. Our premise is the following: in all the examples so far tackled numerically (or analytically) the numerical prediction $(\pi_h, \mathscr{S}_h)$ (or $(\pi, \mathscr{S})$ once an analytical solution of $(\widetilde{\mathscr{P}^*})$ was available) satisfy the property $\operatorname{spt}(\pi_h, \mathscr{S}_h) \subset (\operatorname{spt} f \cup \partial\Omega)^2$. In other words it was enough that the strings connect the points that were either loaded or located on the boundary. If we knew that this is always the case the existence of solution to $(\widetilde{\mathscr{P}^*})$ could be inferred for a discrete load f by exploiting assertion (ii) in Lemma 5.54 (as we shall prove below). Whether this property is true is at the moment an open issue. In what follows we show that it is connected to a geometrical problem of extending monotone maps in a special manner.

Let us be given a monotone multifunction $\mathbf{v}_0$ whose domain $D_0 := \operatorname{dom}(\mathbf{v}_0)$ contains $\mathbb{R}^d \setminus \overline{\Omega}$. Then $\mathbf{v}_0$ admits at least one maximal monotone extension $\mathbf{v}$, that is a maximal monotone map that satisfies $\mathbf{v}(x) \supset \mathbf{v}_0(x)$ for every $x \in \mathbb{R}^d$, see [3]. Owing to the claim (iii) in Proposition 2.39 any such extension is of full domain, i.e. $\operatorname{dom}(\mathbf{v}) = \mathbb{R}^d$. With $\mathbf{v}_0$ we may associate the sub-additive function $c_{\mathbf{v}_0} : D_0 \times D_0 \to \mathbb{R}_+$ given by

$$c_{\mathbf{v}_0}(a, b) := \inf \left\{ \sum_{i=1}^{N-1} \ell_{\mathbf{v}_0}(x_i, x_{i+1}) \ : \ x_1 = a,\ x_N = b, \quad N \geq 2, \quad \{x_i\} \subset D_0 \right\}$$

where $\ell_{\mathbf{v}_0}$ is defined as in (5.122). Note that definition of $c_{\mathbf{v}_0}$ above differs from the original definition in (5.127) since now we may use x_i in the set D_0 only. It is suspected that if $D_0 \cap \Omega$ is finite, among the maximal monotone extensions of $\mathbf{v}_0$ there exists at least one that does not change the induced metric on D_0:

Conjecture 5.106. *Let S be a finite subset of Ω and set $D_0 = S \cup (\mathbb{R}^d \setminus \overline{\Omega})$. Consider a monotone map $\mathbf{v}_0 : D_0 \to \mathbb{R}^d$ such that $\mathbf{v}_0 = \mathrm{id}$ in $\mathbb{R}^d \setminus \overline{\Omega}$.*

Then there exists a maximal monotone extension $\tilde{\mathbf{v}}$ of $\mathbf{v}_0$ such that $c_{\tilde{\mathbf{v}}} = c_{\mathbf{v}_0}$ in $D_0 \times D_0$.

The conjecture is easy to prove in the 1D case: once $\mathbf{v}_0$ is univalued on D_0 the extension $\mathbf{v}$ in question is unique and may be constructed as the interpolation of $\mathbf{v}$ that is affine between each pair of neighbouring points in D_0. One can use assertion (iii) in Proposition 5.100 to prove this. So far the ideas to extend the construction to $d \geq 2$ are rather vague.

The geometric property conjectured above can be translated into the existence of optimal string systems as follows:

Proposition 5.107. *Assume that $S := \mathrm{spt}(f)$ is a finite subset of Ω. Then, if Conjecture 5.106 holds true, there exists a solution $(\pi, \mathscr{S})$ to $(\widetilde{\mathscr{P}^*})$ that satisfies*

$$\mathrm{spt}\,(\pi, \mathscr{S}) \ \subset\ \big(S \times \partial\Omega\big) \cup \big((S \times S) \setminus \Delta\big) \ \subset\ (\mathrm{spt}\, f \cup \partial\Omega)^2.$$

Proof. Since S is finite, the set $G := \big(S \times \partial\Omega\big) \cup \big((S \times S) \setminus \Delta\big)$ is closed and at a positive distance from the diagonal Δ. We may then apply the assertion (ii) of Lemma 5.54 obtaining the existence of an optimal solution for the problem $(\widetilde{\mathscr{P}^*_G})$. Clearly we have $\min \widetilde{\mathscr{P}^*_G} \geq \inf \widetilde{\mathscr{P}^*}$. To show the converse inequality, we use the duality relation

$$\min \widetilde{\mathscr{P}^*_G} = -h^{**}_G(0,0) = -h_G(0,0) = \sup\left\{ \int u\, df \ :\ (u,v) \in \mathscr{B}_G \right\}, \tag{5.161}$$

where h_G is given by (5.92) while, after the change of variable $v = \mathrm{id} - w$, the admissible set associated with $h_G(0,0)$ becomes

$$\mathscr{B}_G := \left\{ (u,v) \in C(\mathbb{R}^d; \mathbb{R}^{d+1}) \ :\ \begin{array}{l} u = 0, \quad v = \mathrm{id} \text{ on } \mathbb{R}^d \setminus \Omega, \\ u(x) - u(y) \leq \ell_v(x,y) \quad \forall (x,y) \in G \end{array} \right\}.$$

Let us set $D_0 := S \cup \mathbb{R}^d \setminus \overline{\Omega}$ and let (u,v) be an element of $\mathscr{B}_G$. Then, denoting by $\mathbf{v}_0 : D_0 \to \mathbb{R}^d$ the monotone map defined by $\mathbf{v}_0(x) = \{v(x)\}$, we obtain $u(x) - u(y) \leq \ell_{\mathbf{v}_0}(x,y)$ for all $(x,y) \in D_0 \times D_0$. Indeed the condition $u(x) - u(y) \leq \ell_v(x,y)$ holding for $(x,y) \in G$ extends to $D_0 \times D_0$ by the symmetry of ℓ_v and the fact that $u = 0$ on $\mathbb{R}^d \setminus \Omega$. Owing to the definition of $c_{\mathbf{v}_0}$ given above (and the fact that $\ell_{\mathbf{v}_0} = \ell_v$ on $D_0 \times D_0$), it is then straightforward that $u(x) - u(y) \leq c_{\mathbf{v}_0}(x,y)$ for all $(x,y) \in D_0 \times D_0$. Following the conjecture for such a $\mathbf{v}_0$ there exists a maximal monotone extension $\tilde{\mathbf{v}} \in \mathbf{V}_\Omega$ such that $c_{\tilde{\mathbf{v}}} = c_{\mathbf{v}_0}$ in $D_0 \times D_0$. Then we define $\tilde{u} : \Omega \to \mathbb{R}$ by

$$\tilde{u}(x) := \inf\Big\{ u(y) + c_{\tilde{\mathbf{v}}}(x,y) \ :\ y \in D_0 \Big\}.$$

By construction $\tilde{u}$ satisfies $\tilde{u}(x) - \tilde{u}(y) \leq c_{\tilde{\mathbf{v}}}(x,y)$ for all $(x,y) \in \overline{\Omega} \times \overline{\Omega}$ while $\tilde{u} = u$ on $D_0 \supset \mathrm{spt}\, f$. It follows that $\tilde{u} \in C_0(\Omega)$ and that $\int \tilde{u}\, df = \int u\, df$. The new pair $(\tilde{u}, \tilde{\mathbf{v}})$ is therefore admissible for the geometric problem $(\mathscr{P}_{\mathrm{geo}})$. Accordingly, thanks to Theorem 5.50, we infer that

$$\inf \widetilde{\mathscr{P}^*} = \mathcal{Z} \geq \int u\, df.$$

The latter inequality being true for all $(u,v) \in \mathscr{B}_G$, we may pass to the supremum with the help of (5.161) thus concluding with the desired inequality $\inf \widetilde{\mathscr{P}^*} \geq \sup\big\{ \int u\, df : (u,v) \in \mathscr{B}_G \big\} = \min \widetilde{\mathscr{P}^*_G}$. □

5.7.2. Optimality conditions in case of non-Lipschitz solutions (u, w)

Throughout Chapter 5 we have derived two types of optimality criteria: the system of PDEs in Theorem 5.41 for the pair of problems $(\overline{\mathscr{P}})$ and $(\mathscr{P}^*)$ and the system based on the two-point relations in Proposition 5.57 for the pair $(\overline{\mathscr{P}})$ and $(\widetilde{\mathscr{P}^*})$. Both results have their limitations. In Theorem 5.41 we assume that the candidates (u, w) for solutions of $(\overline{\mathscr{P}})$ are Lipschitz continuous, meanwhile in general we have to account for the case when functions u is an element of $C^{0,\frac{1}{2}} \cap W^{1,2}(\Omega)$ and w is merely an element of $BV \cap L^\infty(\Omega; \mathbb{R}^d)$, see Theorem 5.25. In order to work with the second result in Proposition 5.57 we only need to enforce that (u, w) are continuous. By using the additional tools borrowed from the theory of maximal monotone maps this restriction was lifted: in Theorem 5.95 any pair $(u, w) \in \overline{\mathcal{X}}_1$ may be considered or more precisely $(u, \mathbf{v}) = \big(u, \mathbf{i}^{-1}(\mathrm{id} - w)\big)$. Both the systems of optimality conditions in Theorem 5.95 and Proposition 5.57, however, have solutions if and only if $(\widetilde{\mathscr{P}^*})$ admits a solution, which may be false, see Remark 5.59.

While for the pair of problems $(\overline{\mathscr{P}})$ and $(\widetilde{\mathscr{P}^*})$ we cannot do better than in Theorem 5.95, there is a room for improving the generality of the system of PDEs in Theorem 5.41. The issue is that the theory of μ-tangential calculus is not directly applicable for functions $(u, w) \in \overline{\mathcal{X}}_1$. Let us stress that the desire to generalize the PDEs is not only to have a more general tool for solving the problems analytically but more importantly to extract the local information about the optimal solutions. For instance, the equation (v) $q = \sigma \nabla_\mu u$ in (5.67) will prove to be fundamental for the main result in Chapter 6. The simplest way to generalize the system of PDEs seems to lead through adapting the theory of μ-tangential calculus in Section 2.5 to functions $(u, w) \in \overline{\mathcal{X}}_1$, or rather independently to u and w. In other words we need to be able to transfer the differential properties of (u, w) known $\mathcal{L}^d$-a.e. in Ω to measures μ that may be lower dimensional. Observe that Remark 5.36 already shows some promise.

Let us start with a function u that together with some w creates a pair in the set $\overline{\mathcal{X}}_1$. In case of function u there is a hope that the μ-tangential theory does not need to be changed – in other words we wish that in Theorem 5.41 we may work with a measure $\mu \in \mathcal{M}_+(\overline{\Omega})$ so that any such function u is an element of $W^{1,2}_\mu$, i.e. $u \in \mathrm{dom}(\nabla_\mu)$ with $q = 2$, cf. Section 2.5.2. Actually, we may adjust the definition of $W^{1,2}_\mu$ by putting $|||\cdot||| := \|\cdot\|_{\infty,\Omega}$ instead of $\|\cdot\|_{L^2_\mu}$, see (2.22). Eventually we ask a question:

Problem 5.108. *Let us take any measure $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ such that $\mathrm{Div}\,\varsigma = 0$ in Ω. Does there exist a representation $\varsigma = \sigma\mu$ with $\mu \in \mathcal{M}_+(\overline{\Omega})$ and $\sigma \in L^1_\mu(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ such that for each pair $(u, w) \in \overline{\mathcal{X}}_1$ the function u is an element of $W^{1,2}_\mu$?*

At the moment this question remains unresolved for arbitrary measures ς. We may, however, point to certain classes of ς for which the problem above has a positive answer. First we make an observation:

Proposition 5.109. *For a measure $\mu \in \mathcal{M}_+(\overline{\Omega})$ assume that $\mathrm{Div}(P_\mu\,\mu) = 0$ in Ω. Then $u \in W^{1,2}_\mu$ whenever $(u, w) \in \overline{\mathcal{X}}_1$ for some w.*

Proof. By definition of $\overline{\mathcal{X}}_1$ we can point to a sequence $\{(u_n, w_n)\} \subset \mathcal{X}_1$ such that $u_n \to u$ uniformly on $\overline{\Omega}$. Since $\mathcal{A}(u_n, w_n) = \frac{1}{2}\nabla u_n \otimes \nabla u_n + e(w_n) \preceq \mathrm{I}$ we have

$$\frac{1}{2}\int_{\overline{\Omega}} |\nabla_\mu u_n|^2 d\mu = \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n, P_\mu\, \mu \right\rangle = \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n + e(w_n), P_\mu\, \mu \right\rangle \leq \mu(\overline{\Omega}).$$

The fact that $u \in W^{1,2}_{\mu,0}(\Omega)$ follows by Corollary 2.55. □

In the context of the work [39] the equation $\mathrm{Div}(P_\mu\, \mu) = 0$ can be called *the zero mean curvature condition for* μ. Below we give two examples of such measures.

Example 5.110. Let A be an affine k-dimensional subspace of $\mathbb{R}^d$ that intersects Ω. Then the measure $\mu_A = \mathcal{H}^k \llcorner (A \cap \Omega)$ satisfies the zero mean curvature condition $\mathrm{Div}(P_{\mu_A}\, \mu_A) = 0$ in Ω and effectively $u \in W^{1,2}_{\mu_A,0}(\Omega)$. This fact is a measure theoretic restatement of $u_A \in W^{1,2}(\Omega_A)$, see Remark 5.36. △

Thanks to example above we learn that Problem 5.108 has an affirmative answer for $\varsigma = \varsigma_{\mathscr{S}}$ visualized in Fig. 5.5(a), (c): it is enough to take $\mu = \mathrm{Tr}\,\varsigma$. The next example allows to make a similar conclusion for other finite systems of strings, see e.g. Fig. 5.5(b) or Fig. 5.7(a).

Example 5.111. Let $\mathscr{S} \in \mathcal{M}_+(\overline{\Omega} \times \overline{\Omega})$ be a finitely supported pairing measure that satisfies (ii) in (5.89). For $\varsigma_{\mathscr{S}}$ being the pre-stress defined in (5.86) let us put $\mu = \mathrm{Tr}\,\varsigma_{\mathscr{S}}$. Then $\mathrm{Div}(P_\mu\, \mu) = 0$ in Ω and, as a consequence of Proposition 5.109, $u \in W^{1,2}_\mu$ once $(u, w) \in \overline{\mathcal{X}}_1$ for some w.

To see this we first realize that $\varsigma_{\mathscr{S}} = \sum_{i=1}^N s_i\, \tau^{x_i,y_i} \otimes \tau^{x_i,y_i}\, \mathcal{H}^1 \llcorner [x_i, y_i]$ for some $s_i > 0$ and $(x_i, y_i) \in \overline{\Omega} \times \overline{\Omega}$. As a result $\mu = \sum_{i=1}^N \mu_i$ where $\mu_i = s_i \mathcal{H}^1 \llcorner [x_i, y_i]$. According to Proposition 2.59 we have $P_\mu = \tau^{x_i,y_i} \otimes \tau^{x_i,y_i}$ μ_i-a.e. We therefore deduce that $P_\mu = \sigma := \frac{d\varsigma_{\mathscr{S}}}{d\mu}$ and eventually $\mathrm{Div}(P_\mu\, \mu) = \mathrm{Div}\,\varsigma_{\mathscr{S}} = 0$ in Ω by (5.89). △

Sadly it is not true in general that $\mathrm{Div}(P_\mu\, \mu) = 0$ when $\mu = \mathrm{Tr}\,\varsigma$ for ς satisfying $\mathrm{Div}\,\varsigma = 0$, even if $\varsigma = \varsigma_{\mathscr{S}}$ for some $\mathscr{S}$. In 2D for instance, the reason lies in the two-dimensional regions where $\sigma = \frac{d\varsigma}{d\mu}$ is rank-one – then $P_\mu = \mathrm{I} \neq \sigma$ in these regions. Nonetheless, such measures are not a lost cause after all:

Example 5.112. For two-dimensional ambient space, i.e. for $d = 2$, take a measure $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ satisfying $\mathrm{Div}\,\varsigma = 0$ in Ω. Let $\varsigma = \sigma_a\, \mathcal{L}^2 \llcorner \Omega + \varsigma_s$ be the Lebesgue decomposition of ς with respect to Lebesgue measure, in particular $\varsigma_s \perp \mathcal{L}^2 \llcorner \Omega$. Assume that $\mu_s := \mathrm{Tr}\,\varsigma_s$ is of the form $\mu_s = \sum_{i=1}^N s_i\, \mathcal{H}^1 \llcorner S_i$ where S_i are C^2 curves. Let us fix $\mu := \mathcal{L}^2 \llcorner \Omega + \mu_s$. Then any u such that $(u, w) \in \overline{\mathcal{X}}_1$ (for some w) is an element of $W^{1,2}_{\mu,0}(\Omega)$.

We may show this by considering a sequence (u_n, w_n) as in the proof of Proposition 5.109. Arguing as in Example 5.111 we may deduce that μ_s-a.e. we have $P_\mu = \sigma_s := \frac{d\varsigma}{d\mu_s}$. Then:

$$\begin{aligned}
&\frac{1}{2}\int_{\overline{\Omega}} |\nabla_\mu u_n|^2 d\mu = \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n, P_\mu\, \mu \right\rangle = \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n, \mathrm{I}\,\mathcal{L}^2 + \sigma_s \mu_s \right\rangle \\
&= \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n, \varsigma_s \right\rangle + \tfrac{1}{2}\left\|\nabla u_n\right\|_{L^2(\Omega)} \leq \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n, \sigma_a \mathcal{L}^2 + \varsigma_s \right\rangle + C \\
&= \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n, \varsigma \right\rangle + C = \int_{\overline{\Omega}} \left\langle \tfrac{1}{2}\nabla u_n \otimes \nabla u_n + e(w_n), \varsigma \right\rangle + C \leq \int_{\overline{\Omega}} \operatorname{Tr}\varsigma + C
\end{aligned}$$

where the constant $C > 0$ is chosen in accordance with the estimate (iii) in Lemma 5.24. Our claim follows. △

It seems difficult to resolve Problem 5.108 in general case and modification of theory of the measure-tangential differential operator summarized in Section 2.5 may turn out unavoidable in the end. From the arguments made above we can see that once $\operatorname{Div}\varsigma = 0$ in Ω we can provide a uniform upper bound for the integral $\int \langle \nabla u \otimes \nabla u, \varsigma\rangle$ for $(u, w) \in \mathcal{X}_1$. Meanwhile estimating $\|\nabla_\mu u\|^2_{L^2_\mu} = \int \langle \nabla u \otimes \nabla u, P_\mu\, \mu\rangle$ causes trouble. This observation encourages to define a ς-tangential operator ∇_ς and the related Sobolev space $W^{1,2}_\varsigma$. Both the new notions would be anisotropic by nature. One can imagine that inclusion $W^{1,2}_\mu \subset W^{1,2}_\varsigma$ would hold true for $\mu = \operatorname{Tr}\varsigma$ since $\operatorname{Im}\sigma \subset \operatorname{Im} P_\mu$ for $\sigma = \frac{d\varsigma}{d\mu}$ by assertion (i) in Theorem 2.57 and Proposition 2.66. This matter was not examined in detail and thus we may pose another problem:

Problem 5.113. *For $\varsigma \in \mathcal{M}_+(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ how to develop the theory of ς-tangential calculus in the form of operator ∇_ς and the Sobolev space $W^{1,2}_\varsigma$?*

In the case of the function $w \in \mathcal{W}_\Omega$ developing the μ-tangential (or ς-tangential) apparatus becomes a much greater challenge. Since w is merely a BV function it is not *a priori* well defined μ-a.e. if μ is not absolutely continuous with respect to Lebesgue measure. Remark 5.36 teaches us that we should aim for working with a μ-tangential function $w_\mu \in L^\infty_\mu(\overline{\Omega}; T_\mu)$ that for a smooth function w would be computed as $w_\mu = P_\mu w$. Simple analysis shows that the orthogonal component $w_\perp = w - w_\mu$ cannot be determined when we pass to arbitrary $w \in \mathcal{W}_\Omega$. Next we must come up with some sort of μ-tangential space $BV_\mu(\overline{\Omega}; T_\mu)$ and the related operator $\varepsilon_\mu = \varepsilon_\mu(w_\mu) \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$. A similar effort was made in [13] but the results obtained therein are far from being directly applicable here. First of all, working with ς instead of μ may be necessary, see Problem 5.113. At the present time, in case of function $w \in \mathcal{W}_\Omega$ the author faces too many issues to formulate crisp ideas. We shall merely pose a very general question:

Problem 5.114. *For $\varsigma \in \mathcal{M}_+(\overline{\Omega}; \mathcal{S}^{d\times d}_+)$ how to develop the theory of ς-tangential calculus in the form of operator ε_ς and the space BV_ς that are adhering to functions $w \in \mathcal{W}_\Omega$?*

5.7.3. The supporting zone Υ_0 other than $\partial\Omega$

From the very beginning of this chapter the membrane was fixed – both in plane and out of plane – on the whole boundary $\partial\Omega$. In the problem $(\mathscr{P})$ it was reflected

by the Dirichlet condition $(u, w) = (0, 0)$ on $\partial\Omega$, whilst in problem $(\mathscr{P}^*)$ it resulted in equilibrium equations $-\operatorname{div}\vartheta = f$ and $\operatorname{Div}\varsigma = 0$ enforced in the interior Ω only. A natural generalization of the optimal membrane problem would be to account for supporting zones (or Dirichlet zones) Υ_0 being arbitrary compact subsets of $\overline{\Omega}$, contained in $\partial\Omega$ or not. This approach is more consistent with respect to Chapter 4 on the Free Material Design problem.

We shall now shortly discuss what changes would the theory of optimal membranes undergo for such a general Dirichlet zone Υ_0. As already implied, in the problem $(\mathscr{P}^*)$ the two constraints will now read: $-\operatorname{div}\vartheta = f$ in $\mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0)$ and $\operatorname{Div}\varsigma = 0$ in $\mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$. The natural question concerns the existence of solutions of $(\mathscr{P}^*)$, which for $\Upsilon_0 = \partial\Omega$ was assured for any $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$. The necessary condition for the existence may be quickly deduced from the result stated in [45, Theorem 2.2]:

Theorem 5.115 (**Bouchitté-Mattei-Milton-Seppecher**). *Let us consider a measure $\varsigma \in \mathcal{M}^b(\mathbb{R}^d; \mathcal{S}^{d\times d}_+)$, i.e. ς admits values in the space of positive semi-definite matrices. For a compact set Υ_0 assume that $\operatorname{Div}\varsigma = 0$ in $\mathcal{D}'(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^d)$. Then*

$$\operatorname{spt}\varsigma \subset \operatorname{co}(\Upsilon_0).$$

Since necessarily $\vartheta \ll \operatorname{Tr}\varsigma$ in order that $\int \varrho^0(\vartheta, \varsigma) < \infty$, from the result above we infer that also $\operatorname{spt}\vartheta \subset \operatorname{co}(\Upsilon_0)$. Eventually we learn that the load f must be supported in $\operatorname{co}(\Upsilon_0)$ as well. It is thus reasonable to demand that:

$$\overline{\Omega} \subset \operatorname{co}(\Upsilon_0). \tag{5.162}$$

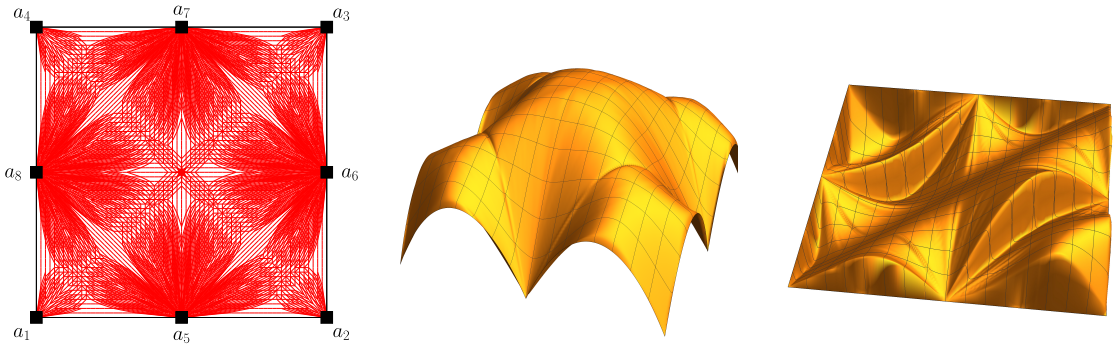
By virtue of the claim (i) in Lemma 5.12 the functional $\mathcal{M}(\overline{\Omega}; \mathbb{R}^d \times \mathcal{S}^{d\times d}_+) \ni (\vartheta, \varsigma) \mapsto \int \varrho^0(\vartheta, \varsigma) = \int \chi^*_{\mathrm{C}}(\vartheta, \varsigma)$ is coercive therefore in order to show existence for problem $(\mathscr{P}^*)$ for arbitrary Dirichlet zone Υ_0 one only needs to point to an admissible pair (ϑ, ς) such that $\int \varrho^0(\vartheta, \varsigma) < \infty$. Such pair may be constructed explicitly (we omit to describe the construction here) and therefore existence of solution to $(\mathscr{P}^*)$ follows.

The differences are much more substantial in the case of problem $(\mathscr{P})$ or, more accurately, its relaxation $(\overline{\mathscr{P}})$. For general Υ_0 satisfying (5.162) the estimates stated in Lemma 5.24 fail to hold true, in particular we cannot expect that functions $w \in \mathcal{X}_1$ are uniformly bounded – they may blow up e.g. on edges of polygon Ω if Υ_0 consists of the vertices only. At this moment generalization of Lemma 5.24 and effectively the relaxation $(\overline{\mathscr{P}})$ for $\Upsilon_0 \neq \partial\Omega$ has not been studied to greater extent. Let us leave this matter here as another open problem:

Problem 5.116. *How to relax the problem $(\mathscr{P})$ in case of the Dirichlet zone being any compact set $\Upsilon_0 \subset \overline{\Omega}$?*

The delicate issues outlined above do not stand in the way of implementing the general Dirichlet zone Υ_0 in the numerical method developed in Section 5.5.4. The modification of the algorithm may be summed up as defining $\mathrm{X}^h_0 = \mathrm{X}^h \cap \Upsilon_0$ instead of $\mathrm{X}^h_0 = \mathrm{X}^h \cap \partial\Omega$. The results like Proposition 5.65 or Theorem 5.66 would still require revision, but it does not hurt to run the MATLAB code and analyse the results, which we will do now:

Example 5.117 (**Eight-point Dirichlet zone**). For the square domain $\Omega = (-\frac{1}{2}, \frac{1}{2})^2$ we again consider a uniform pressure load $f = \mathcal{L}^2 \llcorner \Omega$ but this time we set an eight-point Dirichlet zone $\Upsilon_0 = \{a_1, \ldots, a_8\}$, see Fig. 5.14(a). Numerical solutions $\sigma_{\mathscr{S}_h}$, u_h are presented in Fig. 5.14(a,b) respectively while Fig. 5.14(c) shows the component of w_h along diagonal $[a_4, a_2]$, i.e. $\langle w_h, \tau^{a_4,a_2}\rangle$. Similarly as in Example 5.75 (non-convex Ω) we may predict jump-type discontinuities of exact optimal $w \in BV(\Omega; \mathbb{R}^d)$ along segments $[a_5, a_6]$ and $[a_7, a_8]$, more precisely it is the normal component that is discontinuous (the component $\langle w, \tau^{a_3,a_1}\rangle$ is discontinuous along $[a_5, a_8]$ and $[a_6, a_7]$). Notwithstanding this, the graph of u_h shown in Fig. 5.14(b) seems to point to Lipschitz continuity for the exact solution u. △



(a) (b) (c)

Figure 5.14: Numerical prediction of optimal membrane for uniform pressure load $f = \mathcal{L}^2 \llcorner \Omega$ and an eight-point Dirichlet zone Υ_0: (a) optimal $\varsigma_{\mathscr{S}_h}$; (b) optimal u_h; (c) the component $\langle w_h, \tau^{a_4,a_2}\rangle$ of optimal w_h.

5.7.4. Recasting solution of $(\mathscr{P}^*)$ from transport plan optimal for $c_{\mathbf{v}}$-cost

In Section 5.6 an unexpected connection between the optimal membrane problem and the optimal transport problem was put forth. The fact that from the solution of (OPM) we may extract cost function $c_{\overline{\mathbf{v}}}$ that in the certain class maximizes the Monge-Kantorovich cost $W^{\partial\Omega}_{c_{\overline{\mathbf{v}}}}(f_+, f_-)$ is remarkable. However, at the present moment this link does not deliver a consistent method for solving (OPM) problem as it was the case for the scalar (FMD) problem – the latter can be completely tackled by solving the Monge-Kanotrovich problem for the Euclidean cost function, cf. Section 4.8. Below we shortly discuss the obstacles that we must face on the route from solution $\overline{\gamma}$ of $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ to solution $\overline{\varsigma}$ of (OPM). Convexity of Ω will be assumed.

First we reinterpret formula (4.72) that in Section 4.8 gives the heat flow $\vartheta_{\overline{\gamma}}$ that is optimal for $(\mathcal{P}^*)$. Recall that $\vartheta_{\overline{\gamma}}$ decomposed to straight segments called transport rays. It was so since $\overline{\gamma}$ was an optimal plan for the Monge-Kanotorvich problem for Euclidean cost and therefore the corresponding geodesics were straight segments indeed (if Ω is convex, see [35]). If we now wish to construct a partial solution $\vartheta_{\overline{\gamma}}$ of $(\mathscr{P}^*)$ in a similar fashion then we must employ geodesics with respect to the $c_{\overline{\mathbf{v}}}$ pseudo-metric (or $d_{\overline{\mathbf{v}}}$ metric). This way we run into trouble since definition of such a vector valued measure $\vartheta_{\overline{\gamma}}$ requires that

a tangent vector is well defined at $\mathcal{H}^1$-a.e. point of the geodesic curve. Meanwhile, for a fixed pair $(a,b)\in\overline{\Omega}\times\overline{\Omega}$, the claim (ii) in Proposition 5.100 merely guarantees that the $d_{\overline{\mathbf{v}}}$-geodesics $\Theta=\Theta^{a,b}:[0,1]\to X_{\overline{\mathbf{v}}}$ are Lipschitz continuous with respect to $d_{\overline{\mathbf{v}}}$ metric, but not necessarily with respect to Euclidean metric. In particular it may happen that each selection $\theta^{a,b}:[0,1]\to\overline{\Omega}$ of $\Theta^{a,b}$ is discontinuous, cf. Example 5.99. It does not mean, however, that there is no alternative geodesic $\widetilde{\Theta}^{a,b}$ for which there exists a selection $\widetilde{\theta}^{a,b}$ that is not only continuous but also *rectifible* (of finite length) with respect to Euclidean distance, so that derivative $(\widetilde{\theta}^{a,b})'$ exists $\mathcal{H}^1$-a.e. We thus arrive at the first question that is so far unanswered:

Problem 5.118. *Let* $\mathbf{v}$ *be a maximal monotone map in the set* $\mathbf{V}_\Omega$ *where* Ω *is convex. For a fixed pair of points* $(a,b)\in\overline{\Omega}\times\overline{\Omega}$ *let us consider the* $d_{\mathbf{v}}$*-geodesic problem* (5.153).

Does there exist a continuous rectifible curve $\theta^{a,b}_{\mathbf{v}}:[0,1]\to\overline{\Omega}$ *such that the curve* $\Theta^{a,b}_{\mathbf{v}}=[\theta^{a,b}_{\mathbf{v}}]:[0,1]\to X_{\mathbf{v}}$ *is a* $d_{\mathbf{v}}$*-geodesic, namely it solves* (5.153)*?*

Let us go further assuming that the open problem above can be resolved favourably. Then for each pair of points $(a,b)\in\overline{\Omega}\times\overline{\Omega}$ we may define measure $\vartheta^{a,b}_{\mathbf{v}}\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$ that generalizes $\vartheta^{a,b}$ to non-Euclidean metrics $d_{\mathbf{v}}$:

$$\big\langle\phi,\vartheta^{a,b}_{\mathbf{v}}\big\rangle_{(C,\mathcal{M})}:=\int_{[0,1]}\big\langle\phi\big(\theta^{a,b}_{\mathbf{v}}(t)\big),(\theta^{a,b}_{\mathbf{v}})'(t)\big\rangle\,dt\qquad\forall\,\phi\in C(\overline{\Omega};\mathbb{R}^d).$$

In order to integrate the mapping $\overline{\Omega}\times\overline{\Omega}\ni(a,b)\mapsto\vartheta^{a,b}_\gamma\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$ with respect to $\gamma\in\mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$ we have to show its measurability, which creates another issue:

Problem 5.119. *Assume that the answer to Problem 5.118 is affirmative and fix* $\mathbf{v}\in\mathbf{V}_\Omega$*. Can the mapping* $\overline{\Omega}\times\overline{\Omega}\ni(a,b)\mapsto\theta^{a,b}_{\mathbf{v}}$ *be chosen so that the mapping* $\overline{\Omega}\times\overline{\Omega}\ni(a,b)\mapsto\vartheta^{a,b}_\gamma\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$ *is Borel measurable? (for weak-* topology on* $\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$*)*

If the choice of the measurable mapping above is possible then for each transport plan $\gamma\in\mathcal{M}_+(\overline{\Omega}\times\overline{\Omega})$ we may readily define the measure $\vartheta^{\mathbf{v}}_\gamma\in\mathcal{M}(\overline{\Omega};\mathbb{R}^d)$:

$$\big\langle\phi,\vartheta^{\mathbf{v}}_\gamma\big\rangle_{(C,\mathcal{M})}:=-\int_{\overline{\Omega}\times\overline{\Omega}}\left(\int_{\overline{\Omega}}\langle\phi,\vartheta^{x,y}_{\mathbf{v}}\rangle\right)\gamma(dxdy)\qquad\forall\,\phi\in C(\overline{\Omega};\mathbb{R}^d).\tag{5.163}$$

Next, let us recall Corollary 5.97 which states that once a string system $(\overline{\pi},\overline{\mathscr{S}})$ solves $(\widetilde{\mathscr{P}^*})$ with $\overline{\pi}\le 0$ then $\overline{\gamma}=-\overline{\pi}$ not only solves $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ for any optimal $\overline{v}$ but together with that optimal map also creates a saddle point for the Lagrangian 5.138. This is a strong indication that in order to go from $\overline{\gamma}$ to $\overline{\varsigma}$ solving (OPM) we must consider those solutions $\overline{\gamma}$ of $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ which enjoy the saddle point property. This approach proved to be effective in the case of a one-force load, see observation (3) in Example 5.101. The true challenge arises, however, when the optimal string system in $(\widetilde{\mathscr{P}^*})$ does not exist and we must point to a solution $(\overline{\vartheta},\overline{\varsigma})$ of $(\mathscr{P}^*)$ directly, while we no longer have $c_{\overline{\mathbf{v}}}=\ell_{\overline{\mathbf{v}}}$ $\overline{\gamma}$-a.e. The present author suspects that the construction of $\overline{\vartheta}$ alone may rely on the formula (5.163) similarly as in Section 4.8:

Problem 5.120. *Assume the affirmative answer to Problem 5.119 and take any pair* $(\overline{\gamma},\overline{\mathbf{v}})$ *being a saddle point for the Lagrangian* (5.138) *with* $\nu_1-\nu_2=f$*. Let us put* $\overline{\vartheta}=\vartheta^{\overline{\mathbf{v}}}_{\overline{\gamma}}$*.*

Does there exist $\overline{\varsigma}\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{d\times d}_+)$ *satisfying* $\operatorname{Div}\overline{\varsigma}=0$ *in* Ω *such that* $(\overline{\vartheta},\overline{\varsigma})$ *solves* $(\mathscr{P}^*)$*?*

Assuming that the answer to the problem above is positive we would like to complete the construction and compute the other part of the solution to $(\mathscr{P}^*)$ based on $\overline{\gamma}$: the pre-stress $\overline{\varsigma}$. In Examples 5.101 and 5.103 we have succeeded in doing so (on the level of the string system $\overline{\mathscr{S}}$), whilst Example 5.102 exposed yet another issue: the pre-stress $\overline{\varsigma}$ cannot be locally computed from $\overline{\gamma}$ in regions where $\mathbf{v}$ is tangentially flat, in particular where $\varepsilon\big(\mathbf{i}(\mathbf{v})\big)$ is zero. Without further dwelling on the issue we pose the last question:

Problem 5.121. *Assume that the answer to Problem 5.120 is affirmative. How to actually reconstruct $\overline{\varsigma}$ based on $\overline{\mathbf{v}}$, $\overline{\gamma}$ and/or $\overline{\vartheta} = \vartheta^{\overline{\mathbf{v}}}_{\overline{\gamma}}$?*

The discussion above illustrates how difficult is the path from transport plan $\overline{\gamma}$ optimal for the problem $(\mathrm{OTP}^{\partial\Omega}_{c_{\overline{\mathbf{v}}}})$ to solution of the optimal membrane problem. To answer the questions stated above a far deeper understanding of the pseudo-metric $c_{\mathbf{v}}$ is essential.

5.7.5. Design of pre-loading on the boundary of an elastic membrane

This closing subsection will in turn have a mechanical interest rather than a mathematical one. We shall revisit the very formulation of optimal elastic membrane problem (OEM). Below definition (5.9) of compliance $\mathscr{C}(\lambda)$ in Section 5.2.1 we have emphasized that our membrane model assumes zero displacements on the boundary $\partial\Omega$: both out of plane ($u = 0$) and in plane ($w = 0$). This way without the out-of-plane load f the membrane of a Hooke field $\lambda = \mathscr{H}\mu$ is not stressed at all. Upon exerting the load f, thanks to the in-plane fix on $\partial\Omega$, the in-plane stress σ (relative) or $\varsigma = \sigma\mu$ (absolute) is induced. Although this is mechanically valid, it is not the common way that one is designing a membrane structure: usually the membrane is pre-stressed, i.e. the in-plane stress is introduced prior to applying the load f. In practice there are two main ways to generate such a pre-stress:

(1) freeing the boundary condition for w and applying an in-plane vector load F on $\partial\Omega$;
(2) introducing a non-homogeneous boundary condition for w, for instance $w = \kappa\,\mathrm{id}$ on $\partial\Omega$ where $\kappa \in \mathbb{R}_+$.

Within the approach (1) it is natural to treat the load F as another design variable next to the Hooke field λ. Formulation of the design problem in case of the approach (2) is less obvious. We shall now show that the optimal design of elastic membranes in version (1) is identical to our original framework set in Section 5.2.2.

While discussing the variation (1) we may let go of particular setting of Canvas-like Material Design (CanMD) that we have focused on, starting from Section 5.2.5. Accordingly, we may again consider any triple $\mathcal{K}, j, c$ satisfying conditions listed at the beginning of Section 5.2.1. For a Hooke field $\lambda \in \mathcal{M}(\overline{\Omega}; \mathcal{K})$ and an in-plane load $F \in \mathcal{D}'(\mathbb{R}^d; \mathbb{R}^d)$ (a vector-valued distribution) that is supported on $\partial\Omega$ the compliance reads:

$$\mathscr{C}_F(\lambda) := \sup\left\{ \int u\,df + F(w) - \int j_+\big(\lambda, \mathcal{A}(u,w)\big) \;:\; (u,w) \in \mathcal{D}\big(\Omega\big) \times \mathcal{D}(\mathbb{R}^d; \mathbb{R}^d) \right\};$$

observe that the Dirichlet condition on w is now gone. The modification of the (OEM) problem along the lines of (1) may be readily put forth:

$$\mathscr{C}^{(1)}_{\min} = \inf_{\substack{\lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K})\\ F\in\mathcal{D}'(\mathbb{R}^d;\mathbb{R}^d)}} \left\{ \mathscr{C}_F(\lambda) \ : \int c(\lambda) \leq C_0, \quad \operatorname{spt} F \subset \partial\Omega \right\} \tag{OEM$^{(1)}$}$$

We can see that the "intensity" of F is not constrained – the lack of compactness rules out using the Ky Fan's result (see the proof of Lemma 5.9). Nonetheless, we may always switch inf and sup with an inequality, hence:

$$\begin{aligned}
\mathscr{C}^{(1)}_{\min} &= \inf_{\substack{\lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda)\leq C_0\\ F\in\mathcal{D}'(\mathbb{R}^d;\mathbb{R}^d),\ \operatorname{spt} F\subset\partial\Omega}} \sup_{\substack{u\in\mathcal{D}(\Omega)\\ w\in\mathcal{D}(\mathbb{R}^d;\mathbb{R}^d)}} \left\{ \int u\,df + F(w) - \int j_+\big(\lambda, \mathcal{A}(u,w)\big) \right\} \\
&\geq \sup_{\substack{u\in\mathcal{D}(\Omega)\\ w\in\mathcal{D}(\mathbb{R}^d;\mathbb{R}^d)}} \inf_{\substack{\lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda)\leq C_0\\ F\in\mathcal{D}'(\mathbb{R}^d;\mathbb{R}^d),\ \operatorname{spt} F\subset\partial\Omega}} \left\{ \int u\,df + F(w) - \int j_+\big(\lambda, \mathcal{A}(u,w)\big) \right\} \\
&= \sup_{\substack{u\in\mathcal{D}(\Omega)\\ w\in\mathcal{D}(\Omega;\mathbb{R}^d)}} \inf_{\lambda\in\mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda)\leq C_0} \left\{ \int u\,df - \int j_+\big(\lambda, \mathcal{A}(u,w)\big) \right\} = \mathscr{C}_{\min}.
\end{aligned}$$

The passage to the third line relies on the fact that the infimum in the second line equals $-\infty$ whenever $\operatorname{spt} w$ touches $\partial\Omega$ (while w on $\mathbb{R}^d \setminus \overline{\Omega}$ is immaterial). In the third line we can already employ Ky Fan's theorem to recover the original formulation (OEM), hence the last equality. With the inequality $\mathscr{C}^{(1)}_{\min} \geq \mathscr{C}_{\min}$ at our disposal, we can try to solve the (OEM$^{(1)}$) problem by pointing to an admissible pair (λ, F) for which $\mathscr{C}_F(\lambda) = \mathscr{C}_{\min}$. This strategy ends with a success that renders (OEM$^{(1)}$) fully equivalent to (OEM):

Proposition 5.122. *For (ϑ, ς) being any solution to $(\mathscr{P}^*)$ let λ be constructed in accordance with Corollary 5.19. Then, the pair*

$$\lambda \in \mathcal{M}(\overline{\Omega};\mathcal{K}), \qquad F = -\left(\frac{2C_0}{\mathcal{Z}}\right)^{\frac{1}{2p-1}} \operatorname{Div}\varsigma \quad \in \quad \mathcal{D}(\mathbb{R}^d;\mathbb{R}^d)$$

solves the problem (OEM$^{(1)}$) *of designing a pre-loaded membrane while $\mathscr{C}^{(1)}_{\min} = \mathscr{C}_{\min} = \frac{2p-1}{2p}\big(\frac{\mathcal{Z}^{2p}}{2C_0}\big)^{\frac{1}{2p-1}}$.*

Outline of the proof. The dual definition of $\mathscr{C}_F(\lambda)$ can be derived similarly to (5.13) – the only difference is that the constrain $\operatorname{Div}(\sigma\mu) = 0$ in $\mathcal{D}'(\Omega;\mathbb{R}^d)$ is now replaced by $-\operatorname{Div}(\sigma\mu) = F$ in $\mathcal{D}'(\mathbb{R}^d;\mathbb{R}^d)$. For $\check{\sigma}$ and $\check{\mu}$ constructed as in the proof of Theorem 5.17 we may then estimate $\mathscr{C}_F(\lambda)$ as in the proof of Theorem 5.17 by taking exactly the same $\check{\sigma}$, $\check{q}$ and $\check{\mu}$. This is possible since F is chosen precisely to fit the divergence constraint for $\check{\sigma}\check{\mu}$, see (5.33) and recall that $\hat{\sigma}\hat{\mu} = \varsigma$. □

As mentioned before, the membrane optimization with the kinematically induced pre-stress in accordance with (2) is difficult and is scheduled for the future research of the author. If the Dirichlet condition $w = \kappa\,\mathrm{id}$ is enforced on $\partial\Omega$ the compliance may be defined as

$$\mathscr{C}_\kappa(\lambda) := \sup\left\{ \int u\,df - \int j_+\big(\lambda, \mathcal{A}(u,\widetilde{w}) + \kappa\mathrm{I}\big) \ : \ (u,\widetilde{w}) \in \mathcal{D}(\Omega;\mathbb{R}^{d+1}) \right\};$$

where a change of variables was performed: $w = \tilde{w} + \kappa\,\mathrm{id}$ for $\tilde{w} \in \mathcal{D}(\Omega;\mathbb{R}^d)$. The modified optimal membrane design problem that is parameterized by $\kappa \in \mathbb{R}_+$ may be proposed as follows:

$$\mathscr{C}^{(2)}_{\min,\kappa} = \inf\left\{\mathscr{C}_\kappa(\lambda) \;:\; \lambda \in \mathcal{M}(\overline{\Omega};\mathcal{K}),\ \int c(\lambda) \le C_0\right\} \tag{OEM$^{(2)}_\kappa$}$$

By performing an analysis of (OEM$^{(2)}_\kappa$) similar to the one done for (OEM) problem we may conclude that this modified problem is well posed and solution λ exists. Let us end this chapter with an open question:

Problem 5.123. *Let us fix $\kappa > 0$. What is the relation between solutions λ and λ_κ of problems* (OEM) *and* (OEM$^{(2)}_\kappa$) *respectively?*

Chapter 6

Optimal design of forms

6.1. Formulation of the form-finding problem

Classically, to design a *form* means to design in the three dimensional space a structure that is concentrated on a two dimensional surface S and that can carry the load without bending effects. In the literature the problems of *form-finding* usually concern either objects that are truly 2D (but possibly curved) – these are shells that work in a pure membrane state – or objects that are 1D – spatial trusses whose bars transfer tension/compression while lying on S. The latter structures are typically referred to as *grid-shells*. Quite often we are looking for a form that undergoes stresses of one sign only. We may for instance exclude tensile stresses if we are to design a *vaulted form* made of masonry or glass. On the other hand compression must be ruled out if one aims for a hanging roof made of canvas. Finally, the load to which the designed structure is to be subject can be either prefixed or, to some extent, dependent on the design: wind or gravitational loads might be examples of the latter type.

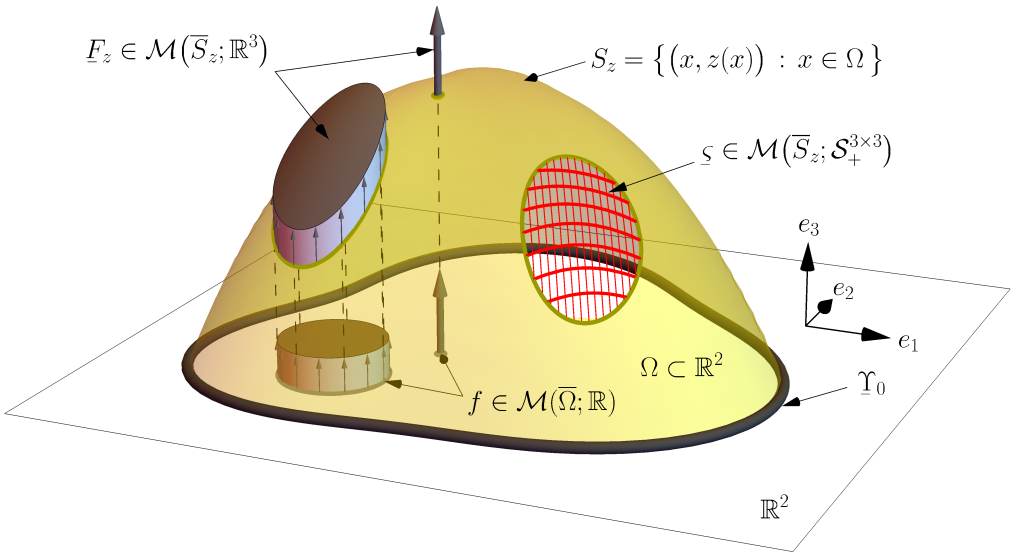


Figure 6.1: The setting of the form-finding problem for $d = 2$, $\underline{d} = 3$.

In this chapter we will address a very particular problem of finding an optimal form. Practically it will fall into the scope of problems described above except that we shall

not *a priori* specify whether our form should be a 2D shell or a grid-shell. Thanks to the measure theoretic approach the dimensionality of the structure is to appear by itself. The data will consist of a plane domain $\Omega \subset \mathbb{R}^d = \mathbb{R}^2$ and the load distribution being a signed measure $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$. Similarly as in the optimal membrane problem the plane is to be treated as horizontal and the scalar field describes the vertical component of the load, meanwhile the horizontal components are assumed to be zero. The formulation of our form-finding problem (in its plastic setting) will be outlined in bullet-points below, see also Fig. 6.1. Let us stress that the under-bar symbol $\underline{\cdot}$ will be henceforward used to emphasize that a given object is three dimensional – either it has values in $\mathbb{R}^{\underline{d}}, \mathcal{S}^{\underline{d}\times\underline{d}}$ where $\underline{d} = d + 1 = 2 + 1 = 3$ or it is embedded in the $\mathbb{R}^{\underline{d}}$ ambient space, or both.

- We are designing an *elevation function* $z \in C_0(\Omega; \mathbb{R})$ (the boundary condition $z = 0$ on $\partial\Omega$ is imposed) whose graph
$$S_z := \Big\{ \underline{z}(x) \,:\, x \in \Omega \Big\}, \qquad \underline{z}(x) := \big(x, z(x)\big)$$
constitutes the surface on which our structure will concentrate.
- The vertical load $\underline{F} = \underline{F}_z$ is then uniquely determined by "lifting" f to the closure $\overline{S}_z$, namely it is defined by means of the push-forward (see Section 2.2):
$$\underline{F}_z := \underline{z} \,\#\, (\mathrm{e}_{\underline{d}} f)$$
where $\mathrm{e}_{\underline{d}}$ is the unit vector that is orthogonal to $\mathbb{R}^d$, namely $\mathbb{R}^{\underline{d}} = \mathbb{R}^d \oplus \mathrm{span}(\mathrm{e}_{\underline{d}})$.
- The curve
$$\Upsilon_0 := \partial\Omega \times \{0\} \;\subset\; \mathbb{R}^{\underline{d}}$$
serves as the Dirichlet region, i.e. as the set on which the structure may rest; note that $\Upsilon_0 = \partial S_z := \overline{S}_z \setminus S_z$.
- Aside from the variable z the other one is a positive semi-definite stress field $\underline{\sigma} \in \mathcal{M}^b(\overline{\Omega} \times \mathbb{R}; \mathcal{S}^{\underline{d}\times\underline{d}}_+)$ that satisfies the equilibrium equation $-\mathrm{Div}\,\underline{\sigma} = \underline{F}_z$ distributionally in $\mathbb{R}^{\underline{d}} \setminus \Upsilon_0$ and moreover
$$\mathrm{spt}\,\underline{\sigma} \;\subset\; \overline{S}_z,$$
i.e. we enforce that $\underline{\sigma} \in \mathcal{M}(\overline{S}_z; \mathcal{S}^{\underline{d}\times\underline{d}}_+)$.
- Among all the pairs $(z, \underline{\sigma})$ satisfying the conditions above we look for the one that minimizes the structural volume understood in the sense of the Michell theory: with $c_{\mathrm{M}} : \mathcal{S}^{\underline{d}\times\underline{d}} \to \mathbb{R}_+$ defined in (4.62) the volume reads
$$\int c_{\mathrm{M}}(\underline{\sigma}) = \frac{1}{\bar{s}_+} \int \mathrm{Tr}\,\underline{\sigma} = \int \mathrm{Tr}\,\underline{\sigma}$$
where we acknowledged that $\underline{\sigma}$ admits values that are positive semi-definite matrices only and we assumed that the yields stress $\bar{s}_+$ is equal to 1.

The *Plastic setting of the Form-Finding problem* may be readily put forward:

$$\inf_{\substack{z \in C_0(\Omega;\mathbb{R}) \\ \underline{\sigma} \in \mathcal{M}^b(\overline{\Omega}\times\mathbb{R}; \mathcal{S}^{\underline{d}\times\underline{d}}_+)}} \left\{ \int \mathrm{Tr}\,\underline{\sigma} \;:\; \mathrm{spt}\,\underline{\sigma} \subset \overline{S}_z, \;\; -\mathrm{Div}\,\underline{\sigma} = \underline{F}_z \;\text{ in } \mathbb{R}^{\underline{d}} \setminus \Upsilon_0 \right\} \tag{PFF}$$

A few words of justification of this design problem are in order. First of all, the vertical load $\underline{F}_z$ is to be interpreted as a gravitational one. Therefore Fig. 6.1, where the load is acting upwards, is misleading: it is more natural to consider negative measures $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R}_-)$ so that the resulting $\underline{F}_z$ is pointing downwards instead. Then there are several scenarios in which the load *vertically migrates* along with the surface S_z or, in other words, it *vertically tracks* this surface. For instance, if S_z plays the role of a roof, the measure $f = -\mathcal{L}^2 \llcorner \Omega$ may represent the snow loading whose intensity in $\mathrm{N/m^2}$ is specified in the design codes with respect to horizontal area – it does not depend on the slope ∇z of the roof. This is not the case for the self-weight load which is dependent both on the geometry of S_z as well as on the material distribution, being a design variable. It is typical, however, that self-weight is neglected in the optimal design: we say that the designed structures are assumed to be *light-weight*. While keeping the interpretation of S_z as a roof we may then think of measure $f = -\delta_{x_0}$ e.g. as a way of modelling the weight of a sculpture that is to be erected on top of the roof at a point whose only the horizontal coordinates are prefixed.

Next we observe that positive semi-definiteness of $\underline{\varsigma}$ is enforced, which means that our form may transfer the load through tensile stresses only, i.e. we are looking for forms made of canvas rather than of masonry. In due course of this chapter (see Corollary 6.19) it will appear that with $f \geq 0$ such optimal forms are to be expected for $z \geq 0$, which agrees with Fig. 6.1. Consequently, if $f \leq 0$ is a gravity loading then the we will have $z \leq 0$, i.e. the optimal form will hang below the horizontal curve $\underline{\Upsilon}_0$. In practice yet another scenario occurs most often: for a gravitational load $f \leq 0$ one is looking to *erect* a form, i.e. to have $z \geq 0$. Such setup requires that the structure is in compression, namely we could enforce that $\underline{\varsigma} \in \mathcal{M}^b\big(\overline{\Omega} \times \mathbb{R}, \mathcal{S}_-^{\underline{d}\times\underline{d}}\big)$ instead. Although such approach is the most attractive from the mechanical point of view, it would be less convenient to handle mathematically. More details on switching between the "pure-tension setting" and the "pure-compression setting" will be given in Section 6.6.1. Until then we will work in the pure-tension setting while the unnatural positive load $f \geq 0$ will be considered in examples in order to have $z \geq 0$, which is simply more attractive in terms of visual outcomes.

Remark 6.1. Let us stress that by keeping the symbols d and $\underline{d} = d+1$ we are automatically accounting for the case when $d = 1$ and $\underline{d} = 2$, i.e. when over a 1D segment $\Omega = (a, b)$ we are looking for a structure that is concentrated on a planar 1D curve (the plane is vertical): a cable (in tension) or an arch (in compression). Nonetheless, we should bear in mind that the case when $d = 2$ and $\underline{d} = 3$ is our main concern.

The plastic formulation (PFF) of the form-finding problem clearly draws upon the formulation of Michell problem (MP), cf. Section 4.7.2. It was a somewhat similar case for the plastic formulation of the optimal membrane problem (OPM). For the both problems (MP) and (OPM) an equivalent elastic formulations of optimal design problems were put forth, namely by employing FibMD and, respectively, CanMD settings of anisotropy optimization. The form-finding will be no exception and, due to the positive semi-definiteness

condition on $\underline{\varsigma}$ in (PFF), we choose Canvas-like Material Design anisotropy setting, i.e.

$$\mathcal{K} = \mathcal{K}_{fib}, \qquad j_+(\underline{H},\underline{\xi}) \ \text{ for } \ j(\underline{H},\underline{\xi}) = \tfrac{1}{2}\big\langle \underline{H}\underline{\xi},\underline{\xi}\big\rangle, \qquad c(\underline{H}) = \operatorname{Tr}\underline{H}. \tag{6.1}$$

Note that we did not use the under-bar symbol for $\mathcal{K}, j, c$ since for those objects the under-bar was used in Chapter 3 for different purpose. Up to cosmetics the definition of compliance of body given by a Hooke field $\underline{\lambda} \in \mathcal{M}^b(\overline{\Omega}\times\mathbb{R};\mathcal{K}_{fib})$ and subject to a load $\underline{F} \in \mathcal{M}^b(\overline{\Omega}\times\mathbb{R};\mathbb{R}^{\underline{d}})$ is identical as in Chapter 4 on the Free Material Design problem:

$$\underline{\mathcal{C}}(\underline{\lambda},\underline{F}) := \sup\left\{ \int \langle \underline{v},\underline{F}\rangle - \int j_+\big(\underline{\lambda},\underline{e}(\underline{v})\big) \ : \ \underline{v}\in\mathcal{D}\big(\mathbb{R}^{\underline{d}}\setminus\Upsilon_0;\mathbb{R}^{\underline{d}}\big)\right\}; \tag{6.2}$$

note that $\underline{e}(\underline{v}) = \frac{1}{2}\big(\nabla\underline{v} + (\nabla\underline{v})^\top\big)$ where the under-bars are used only to emphasize the $\underline{d}$-dimensional ambient space. Aside from the under-bar symbols, in comparison to definition in (4.20) the load $\underline{F}$ now plays the role of the second parameter, while the displacement fields are denoted by $\underline{v}$ instead of $\underline{u}$. The *Elastic setting of the Form-Finding problem* reads as follows:

$$\inf_{\substack{z\in C_0(\Omega;\mathbb{R})\\ \underline{\lambda}\in\mathcal{M}^b(\overline{\Omega}\times\mathbb{R};\mathcal{K}_{fib})}} \left\{ \underline{\mathcal{C}}(\underline{\lambda},\underline{F}_z) \ : \ \operatorname{spt}\underline{\lambda}\subset\overline{S}_z, \ \int \operatorname{Tr}\underline{\lambda} \le C_0\right\} \tag{EFF}$$

The problems (PFF) and (EFF) are the first and the only design problems posed in this thesis whose existence cannot be established through the Direct Method of Calculus of Variations (as opposed to problems (FMD) and (OPM) or (OEM)). First of all the problems are non-convex with respect to the pairs $(z,\underline{\varsigma})$ and, respectively, $(z,\underline{\lambda})$. It is not even clear if the two infima are finite, i.e. whether the set of admissible pairs $(z,\underline{\varsigma})$ for (PFF) is non-empty and if there exists a pair $(z,\underline{\lambda})$ admissible for (EFF) yielding finite compliance $\underline{\mathcal{C}}(\underline{\lambda},\underline{F}_z) < \infty$. It is enough to look into the problems of area-minimizing surfaces (see [94]) to suspect that any problem involving optimization with respect to a surface should be very difficult to analyse. This is precisely why most works on form-finding are of numerical nature: see e.g. [23, 24] for optimizing forms that are shells or [154, 49, 119, 79] for the *Force-Density*-type methods developed for grid-shells.

Notwithstanding the issues pointed out above, the problems (PFF) and (EFF) turn out to be exceptional. The very core of this chapter lies in the newly discovered connection between the foregoing form-finding problems – posed in the $\underline{d}$-dimensional space – and the pair of convex $d = (\underline{d}-1)$-dimensional problems $(\overline{\mathscr{P}})$, $(\mathscr{P}^*)$ derived for the optimal membrane problem in the CanMD setting. In a nut-shell: under mild regularity assumption, the solutions of (PFF), (EFF) will be explicitly constructed based on solutions of $(\overline{\mathscr{P}})$, $(\mathscr{P}^*)$. Although the passage will be non-trivial and will require a certain amount of effort, it is fair to deem the present chapter as much shorter and far less technical than the previous two: on the Free Material Design problem and on the optimal membrane problem. This is somewhat illusive since chronologically the author of this thesis first engaged on the form-finding problem (PFF) which then appeared to reduce to the optimal membrane problem (OPM). The latter one has proved to be of independent interest (both mechanically and mathematically) and it initiated a research conducted alongside Guy Bouchitté – the results made up the paper [29] and Chapter 5 in this thesis. The original intention of the present author for Chapter 5 stands as well: it provides the tools for solving

the optimal form finding problem that is studied in the literature much more extensively than problems of designing membranes.

6.2. Prager problem as a relaxation of the form-finding problem

The most apparent feature of the form finding problems (PFF) or (EFF) that distinguish them from a $\underline{d}$-dimensional Michell problem or, respectively, Free Material Design problem in the CanMD setting is the following: the structure represented by measures $\underline{\varsigma}$ or $\underline{\lambda}$ must be supported on the surface $\overline{S}_z$. However, it is not the only one since in addition we let the load $\underline{F} = \underline{F}_z$ to change with this surface.

In this section we put forth another $\underline{d}$-dimensional design problem that at first glance has nothing to do with form-finding – we let go of the concept of the surface $\overline{S}_z$ being a variable and allow measures $\underline{\varsigma}$ or $\underline{\lambda}$ with an arbitrary support in $\overline{\Omega} \times \mathbb{R}$. What remains of the form-finding framework is the possibility of the load $\underline{F}$ to migrate vertically. This idea is not original and may be traced back to the paper [146] where – to honour the memory of William Prager who initiated this line of research – this design problem was referred to as the *Prager problem*. Let us emphasize that the stresses must still be of one sign: we stick to the condition $\underline{\varsigma} \succeq 0$. The correct mathematical formulation of Prager problem has never been given – the purpose of Section 6.2.1 is to fill this gap by employing the measure-theoretic framework and the concept of *transmissible loads*, see [88]. To the plastic formulation, originally intended by Prager et al., we shall add the elastic formulation using the CanMD anisotropy setting. For both formulations the existence of solutions will be proved.

In 2D, namely for $d = 1$ and $\underline{d} = 2$, the solution of Prager problem – also referred to as *Prager structure* – was proved (at a formal level) to constitute "a form", which in this case means that the structure concentrates on a curve, cf. [145]. Aside from radial symmetric case the analogous result has not been established in three dimensions ($d = 2$ and $\underline{d} = 3$), although the suspicion that 2D forms should solve the 3D Prager problem has been there right from the start of the research. Those suspicions were later supported by the numerical simulations, see [92, 107, 117]. In Section 6.2.2 we make a first step in approving of this hypothesis: the value function of the $\underline{d}$-dimensional Prager problem is estimated from below by means of number $\mathcal{Z}$ that is taken from the $d = (\underline{d} - 1)$-dimensional optimal membrane problem, i.e. $\mathcal{Z} = \max \overline{\mathscr{P}} = \min \mathscr{P}^*$.

6.2.1. Formulation of the Prager problem and the existence result

For a d-dimensional bounded domain $\Omega \subset \mathbb{R}^d$ let us fix a positive height parameter $h_0 \in (0, \infty)$ for which we then define a $\underline{d} = (d + 1)$-dimensional design domain as below:

$$\underline{\Omega} := \Omega \times (-h_0, h_0) \ \subset \ \mathbb{R}^{\underline{d}},$$

naturally its closure reads $\overline{\underline{\Omega}} = \overline{\Omega} \times [-h_0, h_0]$. For the compactness result (see Proposition 6.3) we rule out $h_0 = \infty$, i.e. we cannot take $\underline{\Omega} = \Omega \times \mathbb{R}$, which would have been the most natural setting.

In order to pose the Prager problem correctly the notion of a vertically *transmissible load* introduced in [88] must be brought to the measure theoretic setting. For a signed measure $f \in \mathcal{M}(\overline{\Omega};\mathbb{R})$ (defined on the d-dimensional set $\overline{\Omega}$) let the positive measures $f_+, f_- \in \mathcal{M}_+(\overline{\Omega})$ be the unique Jordan decomposition of f, namely $f = f_+ - f_-$ and $f_+ \perp f_-$ as in (2.4). Then we define two subsets $\mathscr{T}_+(\underline{\Omega}, f)$ and $\mathscr{T}_-(\underline{\Omega}, f)$ of the space $\mathcal{M}(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}})$:

$$\mathscr{T}_+(\underline{\Omega}, f) := \Big\{ \underline{F} = \mathrm{e}_{\underline{d}}\, \underline{f} \ : \ \underline{f} \in \mathcal{M}_+\big(\overline{\underline{\Omega}}\big), \quad \pi_{\mathbb{R}^d} \# \underline{f} = f_+ \Big\},$$
$$\mathscr{T}_-(\underline{\Omega}, f) := \Big\{ \underline{F} = \mathrm{e}_{\underline{d}}\, \underline{f} \ : \ \underline{f} \in \mathcal{M}_+\big(\overline{\underline{\Omega}}\big), \quad \pi_{\mathbb{R}^d} \# \underline{f} = f_- \Big\}$$

where $\pi_{\mathbb{R}^d} : \mathbb{R}^{\underline{d}} \to \mathbb{R}^d$ is the projection onto the base plane $\mathbb{R}^d$, namely we have $\pi_{\mathbb{R}^d}(\underline{x}) = x$ if and only if $\underline{x} = (x,t)$ for some $t \in \mathbb{R}$. Accordingly, $\pi_{\mathbb{R}^d} \# \underline{f}$ is the marginal of $\underline{f}$ on $\overline{\Omega}$ and condition e.g. $\pi_{\mathbb{R}^d} \# \underline{f} = f_+$ may be rewritten as: $\underline{f}\big(B \times [-h_0, h_0]\big) = f_+(B)$ for every Borel set $B \subset \overline{\Omega}$. The set of loads that may be obtained through vertical *transmission of the load* f may be defined as:

$$\mathscr{T}(\underline{\Omega}, f) := \mathscr{T}_+(\underline{\Omega}, f) - \mathscr{T}_-(\underline{\Omega}, f), \tag{6.3}$$

where the difference is intended in the sense of the linear space $\mathcal{M}\big(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}}\big)$.

Let us take any continuous function $z : \overline{\Omega} \to [-h_0, h_0]$ and denote $\underline{z}(x) := \big(x, z(x)\big)$. Then it is easy to check that

$$\underline{F}_z := \underline{z} \,\#\, (\mathrm{e}_{\underline{d}} f) \quad \in \quad \mathscr{T}(\underline{\Omega}, f). \tag{6.4}$$

Indeed, first of all $\underline{F}_z$ is a Radon measure on $\overline{\underline{\Omega}}$ (see Section 2.2) and by linearity of the push-forward we have $\underline{F}_z = \mathrm{e}_{\underline{d}}\,(\underline{z} \,\#\, f_+) - \mathrm{e}_{\underline{d}}\,(\underline{z} \,\#\, f_-)$ while by definition $(\underline{z} \,\#\, f_+)\big(B \times [-h_0, h_0]\big) = f_+\big(\underline{z}^{-1}(B \times [-h_0, h_0])\big) = f_+(B)$ for any Borel set $B \subset \overline{\Omega}$ and a similar result holds true for f_-.

Remark 6.2. Let us explain the difference between the load that is being *transmitted* and the load that *migrates* (the latter notion was used in Section 6.1 on form-finding). We say that f migrates to $\underline{F} \in \mathcal{M}(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}})$ if $\underline{F}$ is precisely of the form $\underline{F}_z$. For instance if $f = \delta_{x_0}$ for some $x_0 \in \overline{\Omega}$ then all loads $\underline{F}$ obtained through migration read $\underline{F} = \mathrm{e}_{\underline{d}}\, \delta_{\underline{x}_0}$ where $\underline{x}_0 = (x_0, t)$ for some $t \in [-h_0, h_0]$. Transmission offers much more freedom: for instance the load $f = \delta_{x_0}$ may be transmitted to $\underline{F} = \mathrm{e}_{\underline{d}}\,\big(p\,\mathcal{H}^1 \llcorner [\underline{a}_0, \underline{b}_0] + \sum_{i=1}^n p_i\, \delta_{\underline{x}_{0,i}}\big)$ provided that $\underline{a}_0, \underline{b}_0, \underline{x}_{0,i} \in \big[(x_0, -h_0), (x_0, h_0)\big]$ and that $p\,|\underline{b}_0 - \underline{a}_0| + \sum_{i=1}^n p_i = 1$ for $p, p_i \in \mathbb{R}_+$. One encounters a similar distinction between transport mapping T in the optimal transport problem of Monge and the more general transport plan γ in the Kantorovich problem, see Section 4.8.

Proposition 6.3. *The set $\mathscr{T}(\underline{\Omega}, f)$ is a non-empty convex weakly-* compact subset of $\mathcal{M}(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}})$.*

Proof. To show non-emptiness it suffices to take $\underline{F} = \underline{F}_z$ for $z = 0$. Convexity and weak-* compactness may be showed for $\mathscr{T}_+(\underline{\Omega}, f)$ only. Convexity is thus straightforward due to linearity of the push-forward. Since $\overline{\underline{\Omega}}$ is compact it is easy to show that the mapping $\underline{f} \mapsto \pi_{\mathbb{R}^d} \# \underline{f}$ is weakly-* continuous, which renders $\mathscr{T}_+(\underline{\Omega}, f)$ weakly-* closed. For any

$\underline{F} \in \mathscr{T}_+(\underline{\Omega}, f)$ we have $|\underline{F}|(\overline{\underline{\Omega}}) = \underline{f}(\overline{\underline{\Omega}}) = f_+(\overline{\underline{\Omega}})$ therefore $\mathscr{T}_+(\underline{\Omega}, f)$ is a bounded subset of $\mathcal{M}(\overline{\underline{\Omega}}; \mathbb{R}^{\underline{d}})$ and ultimately also weakly-* compact by Theorem 2.21. □

With the set of transmitted loads $\mathscr{T}(\underline{\Omega}, f)$ at our disposal we can readily pose the Prager problems. The *Plastic setting of the Prager Problem* will read

$$\mathfrak{Z} := \inf_{\substack{\underline{F}\in\mathcal{M}(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}}) \\ \underline{\varsigma}\in\mathcal{M}(\overline{\underline{\Omega}};\mathcal{S}^{\underline{d}\times\underline{d}}_+)}} \left\{ \int \operatorname{Tr} \underline{\varsigma} \;:\; \underline{F} \in \mathscr{T}(\underline{\Omega}, f), \quad -\operatorname{Div} \underline{\varsigma} = \underline{F} \;\text{ in } \mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0 \right\} \tag{PPP}$$

while the *Elastic setting of the Prager Problem* may be formulated as

$$\mathfrak{C} := \inf_{\substack{\underline{F}\in\mathcal{M}(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}}) \\ \underline{\lambda}\in\mathcal{M}(\overline{\underline{\Omega}};\mathcal{K}_{fib})}} \left\{ \underline{\mathcal{C}}(\underline{\lambda}, \underline{F}) \;:\; \underline{F} \in \mathscr{T}(\underline{\Omega}, f), \quad \int \operatorname{Tr} \underline{\lambda} \le C_0 \right\} \tag{EPP}$$

where $\underline{\mathcal{C}}(\underline{\lambda}, \underline{F})$ is defined in (6.2) while we stick to the CanMD anisotropy setting in accordance with (6.1).

Since both Prager problems consist in minimizing with respect to two variables it is reasonable to think of a two-step design process: first we choose a load $\underline{F}$ and then $\underline{\lambda}$ that is optimal for this load (or $\underline{\varsigma}$ depending on the setting). In other words, in the elastic setting:

$$\mathfrak{C} = \inf_{\underline{F}\in\mathcal{M}(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}})} \left\{ \underline{\mathcal{C}}^{\underline{F}}_{\min} \;:\; \underline{F} \in \mathscr{T}(\underline{\Omega}, f) \right\} \tag{6.5}$$

where for any $\underline{F} \in \mathscr{T}(\underline{\Omega}, f)$ we define

$$\underline{\mathcal{C}}^{\underline{F}}_{\min} := \inf_{\underline{\lambda}\in\mathcal{M}(\overline{\underline{\Omega}};\mathcal{K}_{fib})} \left\{ \underline{\mathcal{C}}(\underline{\lambda}, \underline{F}) \;:\; \int \operatorname{Tr} \underline{\lambda} \le C_0 \right\}. \tag{6.6}$$

The minimization problem in (6.6) is non other than the elasticity framework of the Free Material Design problem studied in Section 4.5.1 and in Chapter 4 as a whole. Load $\underline{F}$ now plays the role of a parameter of the a $\underline{d}$-dimensional problem (F$\underline{\text{M}}$D) = (F$\underline{\text{M}}$D$_{\underline{F}}$). From Section 4.2 we know that for each $\underline{F}$ this design problem may be rewritten as a pair of mutually dual $\underline{d}$-dimensional convex problems $(\underline{\mathcal{P}}) = (\underline{\mathcal{P}}_{\underline{F}})$ and $(\underline{\mathcal{P}}^*) = (\underline{\mathcal{P}}^*_{\underline{F}})$:

$$\sup_{\underline{v}\in\mathcal{D}(\mathbb{R}^{\underline{d}}\setminus\underline{\Upsilon}_0;\mathbb{R}^{\underline{d}})} \left\{ \int \langle \underline{v}, \underline{F} \rangle \;:\; \underline{\rho}_+\big(\underline{e}(\underline{v})\big) \le 1 \;\text{ in } \overline{\underline{\Omega}} \right\} \tag{$\underline{\mathcal{P}}_{\underline{F}}$}$$

$$\inf_{\underline{\varsigma}\in\mathcal{M}(\overline{\underline{\Omega}};\mathcal{S}^{\underline{d}\times\underline{d}})} \left\{ \int \underline{\rho}^0_+(\underline{\varsigma}) \;:\; -\operatorname{Div} \underline{\varsigma} = \underline{F} \;\text{ in } \mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0 \right\} \tag{$\underline{\mathcal{P}}^*_{\underline{F}}$}$$

where the mutually polar functions $\underline{\rho}_+, \underline{\rho}^0_+ : \mathcal{S}^{\underline{d}\times\underline{d}} \to \overline{\mathbb{R}}$ satisfy (note that $p = 2$ considering the form of j_+, cf. (6.1))

$$\frac{1}{2}\big(\underline{\rho}_+(\underline{\xi})\big)^2 = \max_{\underline{H}\in\mathcal{K}_{fib},\ \operatorname{Tr}\underline{H}\le 1} j_+(\underline{H}, \underline{\xi}), \qquad \frac{1}{2}\big(\underline{\rho}^0_+(\underline{\sigma})\big)^2 = \inf_{\underline{H}\in\mathcal{K}_{fib},\ \operatorname{Tr}\underline{H}\le 1} j^*_+(\underline{H}, \underline{\sigma}). \tag{6.7}$$

In particular, owing to Lemma 4.2 we have the equality

$$\underline{\mathcal{C}}^{\underline{F}}_{\min} = \frac{(\underline{Z}_{\underline{F}})^2}{2C_0} \qquad \text{where} \qquad \underline{Z}_{\underline{F}} := \sup \underline{\mathcal{P}}_{\underline{F}} = \inf \underline{\mathcal{P}}^*_{\underline{F}}. \tag{6.8}$$

Since we are in the CanMD anisotropy setting, see (6.1), the formulas for $\underline{\varrho}_+, \underline{\varrho}_+^0$ may be taken from Example 3.28:

$$\underline{\varrho}_+(\underline{\xi}) = \max_{i\in\{1,\dots,\underline{d}\}} \big(\lambda_i(\underline{\xi})\big)_+ = \sup_{\underline{\tau}\in\mathbb{R}^{\underline{d}},\ |\underline{\tau}|\le 1} \big\langle \underline{\xi}, \underline{\tau}\otimes\underline{\tau}\big\rangle, \qquad \underline{\varrho}_+^0(\underline{\sigma}) = \begin{cases} \operatorname{Tr}\underline{\sigma} & \text{if } \underline{\sigma}\in\mathcal{S}_+^{\underline{d}\times\underline{d}}, \\ \infty & \text{if } \underline{\sigma}\notin\mathcal{S}_+^{\underline{d}\times\underline{d}}. \end{cases} \tag{6.9}$$

Accordingly, minimization in $(\underline{\mathcal{P}}_{\underline{F}}^*)$ may run through measures $\underline{\varsigma}\succeq 0$ for which $\int \underline{\varrho}_+^0(\underline{\varsigma}) = \int \operatorname{Tr}\underline{\varsigma}$. This way, similarly to (6.5), we can rewrite the plastic setting of the Prager problem (PPP) as a two-step minimization:

$$\underline{\mathfrak{Z}} = \inf_{\underline{F}\in\mathcal{M}(\overline{\Omega};\mathbb{R}^{\underline{d}})} \Big\{ \underline{Z}_{\underline{F}} \,:\, \underline{F}\in\mathscr{T}(\Omega, f)\Big\}. \tag{6.10}$$

By combining the equalities in (6.5), (6.8) and (6.10) an analogue of Lemma 4.2 can be inferred to connect the plastic and elastic settings of the Prager problem:

$$\underline{\mathfrak{C}} = \frac{\underline{\mathfrak{Z}}^2}{2C_0}. \tag{6.11}$$

Based on the foregoing relations it is straightforward to see that:

Proposition 6.4. *Let $(\underline{\varsigma}, \underline{F})$ (resp. $(\underline{\lambda}, \underline{F})$) be a solution of the problem* (PPP) *(resp.* (EPP)*). Then $\underline{\varsigma}$ (resp. $\underline{\lambda}$) solves the problem $(\underline{\mathcal{P}}_{\underline{F}}^*)$ (resp. the problem* (6.6)*).*

The method developed in Section 4.2, that enabled construction of optimal Hooke field based on solution of $(\mathcal{P}^*)$, now allows to produce a solution of the elastic setting of Prager problem from a solution of its plastic setting:

Proposition 6.5. *Let $(\underline{\varsigma}, \underline{F})$ be a solution of the problem* (PPP) *and set $\underline{\mu} = \operatorname{Tr}\underline{\varsigma}$, $\underline{\sigma} = \frac{d\underline{\varsigma}}{d\underline{\mu}}$. Next, choose any $\underline{\mu}$-measurable selection $\mathscr{H}\in L_{\underline{\mu}}^\infty(\overline{\Omega};\mathcal{K}_{fib})$ of the multifunction $\overline{\Omega}\ni\underline{x}\mapsto \underline{\mathcal{K}}_{1,+}\big(\underline{\sigma}(\underline{x})\big)$ (which always exists) and finally put*

$$\underline{\lambda} = \frac{C_0}{\underline{\mathfrak{Z}}}\, \mathscr{H}\,\underline{\mu}.$$

Then the pair $(\underline{\lambda}, \underline{F})$ solves (EPP).

Remark 6.6. Note that for every $\underline{\sigma}\in\mathcal{S}^{\underline{d}\times\underline{d}}$ the set of optimal Hooke tensor $\underline{\mathcal{K}}_{1,+}(\underline{\sigma})$ is intended in the Canvas-like Material Design anisotropy setting, see Example 3.28 for its characterization.

Proof. By optimality of $(\underline{\varsigma}, \underline{F})$ we have $\underline{Z}_{\underline{F}} = \underline{\mathfrak{Z}}$ while $\underline{\varsigma}$ solves $(\underline{\mathcal{P}}_{\underline{F}}^*)$. Hence, owing to Corollary 4.7, $\underline{\lambda}$ solves the problem (6.6) with $\mathcal{C}_{\min}^{\underline{F}} = \frac{(\underline{Z}_{\underline{F}})^2}{2C_0} = \frac{\underline{\mathfrak{Z}}^2}{2C_0}$. The claim follows through (6.5) and (6.11). □

Let us now move on to show that the two Prager problems (PPP) and (EPP) admit solutions $(\underline{\varsigma}, \underline{F})$ and $(\underline{\lambda}, \underline{F})$, respectively. The main difficulty will consist in proving that $\underline{\mathfrak{Z}} < \infty$ and $\underline{\mathfrak{C}} < \infty$. The challenge comes from the fact that the energy function j_+ does not satisfy the ellipticity condition $(\mathrm{A}_{\mathrm{ellip}})$ (see Section 3.2) which is reflected in the non-finiteness of $\underline{\varrho}_+^0$, which in turn enforces that $\underline{\varsigma}\succeq 0$ in (PPP) (see Section 4.9 that discusses similar issues). Consequently, Proposition 4.29 does not apply here and results of

Section 4.3 cannot be depended on. Ultimately, one could show that any non-zero measure $f \in \mathcal{M}(\overline{\Omega};\mathbb{R})$ can be transmitted to a certain $\underline{F} \in \mathscr{T}(\underline{\Omega}, f)$ so that $\underline{Z}_{\underline{F}} = \infty$ and $\underline{\mathcal{C}}^{\underline{F}}_{\min} = \infty$. Our job is to show that there always exists at least one load $\underline{F} \in \mathscr{T}(\underline{\Omega}, f)$ such that those values are finite after all.

To resolve the issues pointed out above we will exploit the structure of the set of admissible displacement functions in problem $(\mathcal{P}_{\underline{F}})$, namely of the set.

$$\underline{\mathcal{U}}_1 := \Big\{ \underline{v} \in \mathcal{D}(\mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0; \mathbb{R}^{\underline{d}}) \,:\, \underline{\rho}_+\big(\underline{e}(\underline{v})\big) \le 1 \ \text{ in } \overline{\underline{\Omega}} \Big\}. \tag{6.12}$$

Owing to the specific form of $\underline{\rho}_+$, see (6.9), the point-wise constraint above may be rewritten as $\underline{e}(\underline{v}) \preceq \underline{\mathrm{I}}$ in $\overline{\underline{\Omega}}$ where $\underline{\mathrm{I}}$ is the $\underline{d} \times \underline{d}$ identity matrix. Since for each pair of points $(x, y) \in \mathbb{R}^{\underline{d}} \times \mathbb{R}^{\underline{d}}$ we have $\big\langle \underline{v}(\underline{y}) - \underline{v}(\underline{x}), \underline{\tau}^{\underline{x},\underline{y}} \big\rangle = \int_{[\underline{x},\underline{y}]} \big\langle \underline{e}(\underline{v}), \underline{\tau}^{\underline{x},\underline{y}} \otimes \underline{\tau}^{\underline{x},\underline{y}} \big\rangle \, d\mathcal{H}^1$ (cf. the proof of Lemma 5.22) the following implication holds true:

$$\underline{v} \in \underline{\mathcal{U}}_1 \qquad \Rightarrow \qquad \big\langle \underline{v}(\underline{y}) - \underline{v}(\underline{x}), \underline{y} - \underline{x} \big\rangle \le \big|\underline{y} - \underline{x}\big|^2 \qquad \forall\, [\underline{x}, \underline{y}] \subset \overline{\underline{\Omega}}; \tag{6.13}$$

note that above we must enforce that the segment $[\underline{x}, \underline{y}]$ is contained in the $\underline{d}$-dimensional domain $\overline{\underline{\Omega}}$ which is non-convex if $\overline{\Omega} \subset \mathbb{R}^d$ is non-convex.

Lemma 6.7. *For any bounded domain $\Omega \subset \mathbb{R}^d$ and any load $f \in \mathcal{M}(\overline{\Omega};\mathbb{R})$ we have*

$$\underline{\mathfrak{Z}} < \infty, \qquad \underline{\mathfrak{C}} < \infty.$$

Proof. Considering (6.10) and (6.11) it is enough to point to a load $\underline{F}_0 \in \mathscr{T}(\Omega, f)$ such that $\underline{Z}_{\underline{F}_0} < \infty$. Let us fix any height parameter $t_0 \in (0, h_0]$ and for functions $\underline{z}_+(x) := (x, t_0)$, $\underline{z}_-(x) := (x, -t_0)$ whose graphs are parallel to $\overline{\Omega} \times \{0\}$ we define

$$\underline{F}_0 = \underline{z}_+ \# \left(\mathrm{e}_{\underline{d}}\, f_+\right) \,-\, \underline{z}_- \# \left(\mathrm{e}_{\underline{d}}\, f_-\right)$$

that is, loosely speaking, we move the positive part of the load f_+ by t_0 above $\overline{\Omega}$ and the negative part f_- below $\overline{\Omega}$ by the same value. It is straightforward to check that $\underline{F}_0 \in \mathscr{T}(\Omega, f)$. We will show that for each $\underline{v} \in \underline{\mathcal{U}}_1$ we have the estimates:

$$\forall\, x \in \overline{\Omega} \ \text{ it holds that } \begin{cases} \big\langle \underline{v}(x, t_0), \mathrm{e}_{\underline{d}} \big\rangle \le C(\Omega, t_0), \\ -\big\langle \underline{v}(x, -t_0), \mathrm{e}_{\underline{d}} \big\rangle \le C(\Omega, t_0), \end{cases} \qquad C(\Omega, t_0) := \frac{D^2}{4t_0} + t_0 \tag{6.14}$$

where $D = \operatorname{diam}(\Omega)$. If this is the case then clearly for each $\underline{v} \in \underline{\mathcal{U}}_1$ we have

$$\int_{\overline{\underline{\Omega}}} \langle \underline{v}, \underline{F}_0 \rangle = \int_{\overline{\Omega}} \big\langle \underline{v}(x, t_0), \mathrm{e}_{\underline{d}} \big\rangle \, f_+(dx) - \int_{\overline{\Omega}} \big\langle \underline{v}(x, -t_0), \mathrm{e}_{\underline{d}} \big\rangle \, f_-(dx) \,\le\, C(\Omega, t_0)\, |f|(\overline{\Omega})$$

which gives $\underline{Z}_{\underline{F}_0} \le C(\Omega, t_0)\, |f|(\overline{\Omega}) < \infty$ thus furnishing the thesis.

To show (6.14) let us fix $\underline{x} = (x_0, t_0)$. We choose a pair of distinct points $y_1, y_2 \in \partial\Omega$ such that $x_0 \in [y_1, y_2]$ and $[y_1, y_2] \subset \overline{\Omega}$ and then we put $\tau = \frac{y_2 - y_1}{|y_2 - y_1|}$ and $\underline{y}_i = (y_i, 0) \in \underline{\Upsilon}_0$ for $i = 1, 2$. Then clearly $[\underline{x}, \underline{y}_i] \subset \overline{\underline{\Omega}}$ and by (6.13) we have $\big\langle \underline{v}(\underline{x}) - \underline{v}(\underline{y}_i), \underline{x} - \underline{y}_i \big\rangle \le |\underline{x} - \underline{y}_i|^2$. Considering that $\underline{v}(\underline{y}_i) = 0$ and $\underline{x} - \underline{y}_i = (x_0 - y_i) + t_0\, \mathrm{e}_{\underline{d}}$ we deduce that

$$\big\langle \underline{v}(\underline{x}), x_0 - y_i \big\rangle + \big\langle \underline{v}(\underline{x}), \mathrm{e}_{\underline{d}} \big\rangle\, t_0 \,\le\, |x - y_i|^2 + (t_0)^2 \qquad i = 1, 2.$$

Then we see that $\langle \underline{v}(\underline{x}), x_0 - y_1 \rangle = \langle \underline{v}(\underline{x}), \tau \rangle\, |x - y_1|$, meanwhile $\langle \underline{v}(\underline{x}), x_0 - y_2 \rangle = -\langle \underline{v}(\underline{x}), \tau \rangle\, |x - y_2|$. Therefore, if we multiply the inequality for $i = 1$ by $|x - y_2|$ and

the one for $i = 2$ by $|x - y_1|$ and then add them side-wise we obtain after additional simplifications that $\langle \underline{v}(\underline{x}), \mathrm{e}_{\underline{d}} \rangle \leq \frac{|x-y_1||x-y_2|}{t_0} + t_0$ and the top inequality in (6.14) follows since $|x - y_1| + |x - y_2| \leq D$. The bottom inequality is shown identically for $\underline{x} = (x, -t_0)$. □

Theorem 6.8. *For any bounded domain $\Omega \subset \mathbb{R}^d$ and any load $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ there always exist solutions of the Prager problems* (PPP) *and* (EPP).

Proof. The problem (PPP) can be rewritten so that we minimize $\int \underline{\rho}^0_+(\underline{\varsigma})$ while we may choose from $(\underline{\varsigma}, \underline{F}) \in \Sigma := \mathcal{M}(\overline{\Omega}; \mathcal{S}^{\underline{d}\times\underline{d}}) \times \mathscr{T}(\Omega, f)$. By Proposition 6.3 the set $\mathscr{T}(\Omega, f)$ is weakly-* compact while the minimized functional is weakly-* lower semi-continuous and coercive with respect to $\underline{\varsigma}$ (cf. the "moreover part" of Proposition 3.11). Existence of a solution of (PPP) readily follows thanks to Lemma 6.7 and the fact that the constraint $-\mathrm{Div}\,\underline{\varsigma} = \underline{F}$ yields a weakly-* closed subset of Σ.

Existence of (EPP) readily follows by Proposition 6.5 but an independent proof might be of interest. We can see that the admissible set $\{\underline{\lambda} \in \mathcal{M}(\overline{\Omega}; \mathcal{K}_{fib}) \,:\, \int \mathrm{Tr}\,\underline{\lambda} \leq C_0\} \times \mathscr{T}(\Omega, f)$ is weakly-* compact, see Proposition 3.33. For each fixed $\underline{v} \in \mathcal{D}(\mathbb{R}^{\underline{d}} \setminus \Upsilon_0; \mathbb{R}^{\underline{d}})$ the functional $(\underline{\lambda}, \underline{F}) \mapsto \int \langle \underline{v}, \underline{F} \rangle - \int j_+(\underline{\lambda}, \underline{e}(\underline{v}))$ is weakly-* lower semi-continuous (jointly in the two variables) owing to Theorem 3.35. Then the functional $\underline{\mathcal{C}} = \underline{\mathcal{C}}(\underline{\lambda}, \underline{F})$ inherits this property as a point-wise supremum of such functionals. Existence for (EPP) is thus established thanks to Lemma 6.7. □

We have succeeded to show that the convex Prager problems (PPP) and (EPP) are both well posed and now it is time to relate them to the form-finding problems (PFF) and (EFF). Owing to (6.4) we can see that effectively the set of admissible designs is larger in the case of the first pair of problems, more precisely

$$\left\{(\underline{F}_z, \underline{\varsigma}) \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^{\underline{d}} \times \mathcal{S}^{\underline{d}\times\underline{d}}_+) \,:\, z \in C_0(\Omega),\ \mathrm{spt}\,\underline{\varsigma} \subset \overline{S}_z\right\} \subset \mathscr{T}(\Omega, f) \times \mathcal{M}(\overline{\Omega}; \mathcal{S}^{\underline{d}\times\underline{d}}_+) \quad (6.15)$$

$$\left\{(\underline{F}_z, \underline{\lambda}) \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^{\underline{d}} \times \mathcal{K}_{fib}) \,:\, z \in C_0(\Omega),\ \mathrm{spt}\,\underline{\lambda} \subset \overline{S}_z\right\} \subset \mathscr{T}(\Omega, f) \times \mathcal{M}(\overline{\Omega}; \mathcal{K}_{fib}) \quad (6.16)$$

and, as a result

$$\infty \geq \inf(\mathrm{PFF}) \ \geq\ \min(\mathrm{PPP}) = \underline{\mathfrak{Z}}, \qquad \infty \geq \inf(\mathrm{EFF}) \ \geq\ \min(\mathrm{EPP}) = \underline{\mathfrak{C}}.$$

One may be tempted to call the Prager problem (PPP) (resp. (EPP)) a relaxation of the form-finding problem (PFF) (resp. (EFF)): after all we have enlarged the set of admissible variables to arrive at a convex problem that admits a solution. However, such a generalization procedure does not quite adheres to the idea of relaxation in the spirit of modern calculus of variations, see e.g. [111], [80], [71]. Ideally, to comply with such an understanding we would have to make sure that the weak-* closed convex hull of the sets on the left hand sides of inclusions (6.15), (6.16) are equal to the sets on the respective right hand sides and moreover that $\inf(\mathrm{PFF}) = \min(\mathrm{PPP})$ and $\inf(\mathrm{EFF}) = \min(\mathrm{EPP})$.

We shall neglect to examine the aforementioned delicate issues and (possibly) abuse the language by saying that the Prager problems are the relaxed optimal form-finding problems. The justification will be other than through standard methods of calculus of variation: in Section 6.3 we shall (under additional regularity assumptions) construct

solutions $(\underline{F}, \underline{\varsigma})$ and $(\underline{F}, \underline{\lambda})$ to Prager problems (PPP) and, respectively, (EPP) that are actually forms, i.e. $\underline{F} = \underline{F}_z$ and $\operatorname{spt} \underline{\varsigma}$, $\operatorname{spt} \underline{\lambda} \subset \overline{S}_z$ for some elevation function $z \in C_0(\Omega)$.

6.2.2. The lower bounds via the theory of optimal membranes

As advertised at the end of Section 6.1 to construct an optimal form in $\underline{d}$-dimensional region $\overline{\underline{\Omega}} = \overline{\Omega} \times [-h_0, h_0]$ we shall exploit solutions of the optimal membrane problem encoded in the pair $(\overline{\mathscr{P}}), (\mathscr{P}^*)$ (see Chapter 5) posed in the d-dimensional region $\overline{\Omega}$. We must strongly emphasize that a particular anisotropy setting of the optimal membrane problem must be chosen: the Canvas-like Material Design setting (CanMD) that was in fact central for Chapter 5, see Section 5.2.5. Although the optimal form-finding problems (PFF), (EFF) could be linked to $(\overline{\mathscr{P}}), (\mathscr{P}^*)$ directly, it is far more convenient (and ultimately much more insightful, see Theorem 6.18) to make a connection between Prager problems (PPP), (EPP) and $(\overline{\mathscr{P}}), (\mathscr{P}^*)$. This connection starts with lower bounds on the value functions in the Prager problem:

Theorem 6.9. *For any bounded domain Ω and any load $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ we have inequalities:*

$$\underline{\mathfrak{Z}} \geq \sqrt{2}\, \mathcal{Z}, \qquad \underline{\mathfrak{C}} \geq \frac{\mathcal{Z}^2}{C_0}$$

where $\mathcal{Z} = \max \overline{\mathscr{P}} = \min \mathscr{P}^*$ is computed for the optimal membrane problem in the CanMD anisotropy setting posed for data Ω, f.

To prove the inequalities above we will first require several lemmas. We propose to enlarge the set $\mathscr{T}(\underline{\Omega}, f)$ of loads that can be obtain by transmission of f to:

$$\widetilde{\mathscr{T}}(\underline{\Omega}, f) := \Big\{ \underline{F} \in \mathcal{M}(\overline{\underline{\Omega}}; \mathbb{R}^{\underline{d}}) \,:\, \pi_{\mathbb{R}^d} \# \underline{F} = \mathrm{e}_{\underline{d}}\, f \Big\}.$$

Clearly a proper inclusion $\mathscr{T}(\underline{\Omega}, f) \subsetneq \widetilde{\mathscr{T}}(\underline{\Omega}, f)$ holds true. First of all, the set $\widetilde{\mathscr{T}}(\underline{\Omega}, f)$ contains loads $\underline{F}$ that are not vertical, i.e. the values of $\underline{F}$ are not necessarily parallel to $\mathrm{e}_{\underline{d}}$. Secondly, the field $\underline{F}$ may consist of components that "vertically cancel each other out": e.g. for a given $\underline{F}_0 \in \mathscr{T}(\underline{\Omega}, f)$, for any $\underline{P} \in \mathbb{R}^{\underline{d}}$ and $x_0 \in \overline{\Omega}$ one has

$$\underline{F} = \underline{F}_0 + \underline{P}\, \delta_{(x_0, h_1)} - \underline{P}\, \delta_{(x_0, h_2)} \quad \in \quad \widetilde{\mathscr{T}}(\underline{\Omega}, f) \qquad \forall\, h_1, h_2 \in [-h_0, h_0]. \tag{6.17}$$

It is thus straightforward that if we were to replace the set $\mathscr{T}(\underline{\Omega}, f)$ by $\widetilde{\mathscr{T}}(\underline{\Omega}, f)$ in the Prager problems (PPP), (EPP) we would acquire value functions that are not greater, e.g. $\inf_{\underline{F} \in \widetilde{\mathscr{T}}(\underline{\Omega}, f)} \underline{Z}_{\underline{F}} \leq \inf_{\underline{F} \in \mathscr{T}(\underline{\Omega}, f)} \underline{Z}_{\underline{F}} = \underline{\mathfrak{Z}}$. For such a modified Prager problem, however, the existence of solution would no longer be clear since $\widetilde{\mathscr{T}}(\underline{\Omega}, f)$ is not weakly-* compact. More importantly, designing the load by choosing from the set $\widetilde{\mathscr{T}}(\underline{\Omega}, f)$ looses physical meaning that was assigned to the migrating load (see Section 6.1) and that – to some extent – was sustained for loads in the set $\mathscr{T}(\underline{\Omega}, f)$. The larger load set will be herein employed only to prove the inequalities in Theorem 6.9 as it will unravel that $\inf_{\underline{F} \in \widetilde{\mathscr{T}}(\underline{\Omega}, f)} \underline{Z}_{\underline{F}} \geq \sqrt{2}\, \mathcal{Z}$ as well, which will be easier to show directly.

We observe that if the number $\underline{Z}_{\underline{F}}$ is expressed as $\sup \underline{\mathcal{P}}_{\underline{F}}$ then in both formulation (PPP), (EPP) the load $\underline{F}$ enters through its virtual work $\int_{\overline{\underline{\Omega}}} \langle \underline{v}, \underline{F} \rangle$. While minimization of

this integral with respect to $\underline{F} \in \mathscr{T}(\underline{\Omega}, f)$ is more involved we learn that by minimizing over the larger set $\widetilde{\mathscr{T}}(\underline{\Omega}, f)$ we test whether $\underline{v}$ varies in the $\underline{d}$-th coordinate:

Lemma 6.10. *For any $\underline{v} \in C(\overline{\underline{\Omega}}; \mathbb{R}^{\underline{d}})$ there holds*

$$\inf_{\underline{F} \in \widetilde{\mathscr{T}}(\underline{\Omega}, f)} \int_{\overline{\underline{\Omega}}} \langle \underline{v}, \underline{F} \rangle = \begin{cases} \int_{\overline{\Omega}} \langle v, \mathrm{e}_{\underline{d}} \rangle \, df & \text{if} \quad \underline{v}(\underline{x}) = v\big(\pi_{\mathbb{R}^d}(\underline{x})\big) \quad \text{for some} \quad v \in C(\overline{\Omega}; \mathbb{R}^{\underline{d}}), \\ -\infty & \text{otherwise,} \end{cases}$$

i.e. the infimum is infinite unless $\underline{v}$ is independent of the $\underline{d}$-th coordinate.

Proof. Assume that there exist distinct points $\underline{x}_1 = (x_0, h_1)$, $\underline{x}_2 = (x_0, h_2) \in \overline{\underline{\Omega}}$ such that $\underline{v}(\underline{x}_1) \neq \underline{v}(\underline{x}_2)$. Then, for any $\underline{F}_0 \in \mathscr{T}(\underline{\Omega}, f)$ and $\underline{P} = -t\big(\underline{v}(\underline{x}_1) - \underline{v}(\underline{x}_2)\big)$ with $t \geq 0$ the load $\underline{F}$ according to (6.17) is an element of $\widetilde{\mathscr{T}}(\underline{\Omega}, f)$ for which $\int_{\overline{\underline{\Omega}}} \langle \underline{v}, \underline{F} \rangle = \int_{\overline{\underline{\Omega}}} \langle \underline{v}, \underline{F}_0 \rangle - t\,|\underline{v}(\underline{x}_1) - \underline{v}(\underline{x}_2)|^2$. By sending t to ∞ we obtain $\inf_{\underline{F} \in \widetilde{\mathscr{T}}(\underline{\Omega}, f)} \int_{\overline{\underline{\Omega}}} \langle \underline{v}, \underline{F} \rangle = -\infty$. If we assume that $\underline{v}$ does not depend on the $\underline{d}$-th coordinate, i.e. that $\underline{v}(\underline{x}) = v\big(\pi_{\mathbb{R}^d}(\underline{x})\big)$, then $\int_{\overline{\underline{\Omega}}} \langle \underline{v}, \underline{F} \rangle = \int_{\overline{\Omega}} \langle v, \pi_{\mathbb{R}^d} \# \underline{F} \rangle$. The latter integral is equal to $\int_{\overline{\Omega}} \langle v, \mathrm{e}_{\underline{d}} \rangle \, df$ for any $\underline{F}$ belonging to $\widetilde{\mathscr{T}}(\underline{\Omega}, f)$ by the very definition. □

Remark 6.11. To some extent the assertion of Lemma 6.10 is not new. In papers [146, 88, 56, 122] (to name a few) the authors already realized that minimizing with respect to transmissible loads leads to the component $\langle \underline{v}, \mathrm{e}_{\underline{d}} \rangle$ (and only this component!) being independent of the vertical coordinate. In those works, however, the load $\underline{F}$ is said to be obtained by transmission if it belongs to the set $\tilde{\mathscr{T}}(\underline{\Omega}, f) := \big\{ \mathrm{e}_{\underline{d}} \, \underline{f} \, : \, \underline{f} \in \mathcal{M}(\overline{\underline{\Omega}}; \mathbb{R}), \ \pi_{\mathbb{R}^d} \# \underline{f} = f \big\}$ (the authors limited their study to discrete loads thus avoiding measure theoretic definitions). This set is also non-compact as it allows self-cancellation although only vertical loads are now in play: eventually we have $\mathscr{T}(\underline{\Omega}, f) \subsetneq \tilde{\mathscr{T}}(\underline{\Omega}, f) \subsetneq \widetilde{\mathscr{T}}(\underline{\Omega}, f)$. Going from the left to the right side of this chain of inclusions we increasingly depart from the physical meaning originally intended for the migrating loads.

Because of Lemma 6.10 we shall now be paying attention to displacement fields $\underline{v} \in C(\overline{\underline{\Omega}}; \mathbb{R}^{\underline{d}})$ that do not depend on the $\underline{d}$-th coordinate. For such functions there always exist functions $w \in C(\overline{\Omega}; \mathbb{R}^d)$ and $u \in C(\overline{\Omega}; \mathbb{R})$ such that (see Remark 6.12 below for the comment on the multiplier $\sqrt{2}$)

$$\underline{v}(\underline{x}) = w(x) + \sqrt{2}\, u(x)\, \mathrm{e}_{\underline{d}} \qquad \forall\, x \in \overline{\Omega} \tag{6.18}$$

where we agree on two conventions that will be henceforward used:

- if $\underline{x}$ and x are present on two respective sides of an equality we understand that it holds for every $\underline{x} = (x, t)$ where $t \in [-h_0, h_0]$;
- a d-dimensional vector, e.g. $w(x) \in \mathbb{R}^d$ is naturally embedded into $\mathbb{R}^{\underline{d}} = \mathbb{R}^d \oplus \mathrm{span}(\mathrm{e}_{\underline{d}})$.

The latter convention will be extended towards embedding matrices $M \in \mathbb{R}^{d \times d}$ or $M \in \mathcal{S}^{d \times d}$ into the space $\mathbb{R}^{\underline{d} \times \underline{d}}$ or, respectively, $\mathcal{S}^{\underline{d} \times \underline{d}}$. The elements of $\mathbb{R}^{\underline{d} \times \underline{d}}$ or $\mathcal{S}^{\underline{d} \times \underline{d}}$ will be often written in the form of *block matrices* in which we agree that elements of $\mathbb{R}^d$ will be always

meant as column vectors. Namely, for elements $M \in \mathbb{R}^{d\times d}$, $a, b \in \mathbb{R}^d$ and $c \in \mathbb{R}$ we will write

$$\begin{bmatrix} M & a \\ b^\top & c \end{bmatrix} := M + a \otimes \mathrm{e}_{\underline{d}} + \mathrm{e}_{\underline{d}} \otimes b + c\, \mathrm{e}_{\underline{d}} \otimes \mathrm{e}_{\underline{d}} \quad \in \quad \mathbb{R}^{\underline{d}\times\underline{d}}.$$

Readily we may compute the gradient of the function $\underline{v}$ that is of the form (6.18) as $\underline{\nabla}\underline{v}(\underline{x}) = \nabla w(x) + \sqrt{2}\,\nabla u(x) \otimes \mathrm{e}_{\underline{d}}$. The symmetric part of this gradient reads

$$\underline{e}(\underline{v})(\underline{x}) = e(w)(x) + \sqrt{2}\,\nabla u(x) \overset{s}{\otimes} \mathrm{e}_{\underline{d}} = \begin{bmatrix} e(w)(x) & \frac{1}{\sqrt{2}}\nabla u(x) \\ \frac{1}{\sqrt{2}}\big(\nabla u(x)\big)^\top & 0 \end{bmatrix}. \tag{6.19}$$

Remark 6.12. The multiplier $\sqrt{2}$ is introduced in (6.18) artificially, i.e. it has nothing to do with the function $\underline{v}$ being independent of $\underline{d}$-th coordinate. It is essential for making the connection to the optimal membrane problem. The $\sqrt{2}$ has already appeared in Theorem 6.9 and will reoccur throughout the remainder of this chapter.

The next result appeals to the particular form of $\underline{e}(\underline{v})$ in (6.19). It is completely new and may be deemed the key for the passage between the Prager problem (and effectively the optimal form-finding problem) and the optimal membrane problem. Let us recall that $\varrho : \mathbb{R}^d \times \mathcal{S}^{d\times d} \to \mathbb{R}$ is the convex one-homogeneous continuous function that yields the equivalence: $\varrho(\theta,\xi) \le 1 \;\Leftrightarrow\; \rho_+(\frac{1}{2}\,\theta\otimes\theta + \xi) \le 1$, see Section 5.2.3. Since we are considering the CanMD setting of the optimal membrane problem the function ρ_+ is the one given in (5.38), i.e. it is the d-dimensional counterpart of $\underline{\rho}_+$ and therefore $\rho_+(\frac{1}{2}\,\theta\otimes\theta + \xi) \le 1$ is in turn equivalent to $\frac{1}{2}\,\theta\otimes\theta + \xi \preceq \mathrm{I}$.

Lemma 6.13. *For a pair $\xi \in \mathcal{S}^{d\times d}$, $\theta \in \mathbb{R}^d$ consider a matrix $\underline{\xi} = \xi + \sqrt{2}\,\theta \overset{s}{\otimes} \mathrm{e}_{\underline{d}}$. Then we have an equality*

$$\underline{\rho}_+(\underline{\xi}) = \varrho\big(\theta,\xi\big).$$

Proof. Both the mappings $\mathbb{R}^d \times \mathcal{S}^{d\times d} \ni (\theta,\xi) \mapsto \varrho\big(\theta,\xi\big)$ and $\mathbb{R}^d \times \mathcal{S}^{d\times d} \ni (\theta,\xi) \mapsto \underline{\rho}_+\big(\xi + \sqrt{2}\,\theta \overset{s}{\otimes} \mathrm{e}_{\underline{d}}\big)$ are convex, lower semi-continuous and positively one-homogeneous. Therefore the assertion will follow if we prove the equivalence: $\underline{\rho}_+(\underline{\xi}) \le 1 \;\Leftrightarrow\; \varrho\big(\theta,\xi\big) \le 1$. To express the condition $\underline{\rho}_+(\underline{\xi}) \le 1$ we will use the latter formula in (6.9): we must test the value of $\big\langle \underline{\xi}, \underline{\tau}\otimes\underline{\tau}\big\rangle$ with vectors $\underline{\tau}$ chosen from the unit sphere $S^{\underline{d}-1}$. Due to the structure of $\underline{\xi}$ we may upfront exclude vectors $\underline{\tau}$ that are parallel to $\mathrm{e}_{\underline{d}}$. The remaining vectors may be written in terms of $\tau \in S^{d-1}$ and the slope $\psi \in \mathbb{R}$ with respect to the base plane $\mathbb{R}^d$:

$$\underline{\tau} = \frac{1}{\sqrt{1+\psi^2}}\,\tau + \frac{\psi}{\sqrt{1+\psi^2}}\,\mathrm{e}_{\underline{d}}$$

and thus

$$\underline{\rho}_+(\underline{\xi}) \le 1 \qquad \Leftrightarrow \qquad \sup_{\tau\in S^{d-1}} \sup_{\psi\in\mathbb{R}} \frac{\langle \xi, \tau\otimes\tau\rangle + \sqrt{2}\,\psi\,\langle\theta,\tau\rangle}{1+\psi^2} \le 1.$$

The condition on the right hand side may be rewritten as

$$\langle \xi, \tau\otimes\tau\rangle + \sqrt{2}\,\psi\,\langle\theta,\tau\rangle \le 1+\psi^2 \qquad \forall\,\psi\in\mathbb{R},\ \tau\in S^{d-1}$$

or

$$\psi^2 - \big(\sqrt{2}\,\langle\theta,\tau\rangle\big)\,\psi + \big(1 - \langle\xi,\tau\otimes\tau\rangle\big) \ge 0 \qquad \forall\,\psi\in\mathbb{R},\ \tau\in S^{d-1}.$$

For a fixed $\tau \in S^{d-1}$ the left hand side of the latter inequality is a univariate quadratic function with respect to ψ. It is non-negative for each ψ only if its discriminant is non-positive, i.e. we reach yet another equivalent condition:

$$2\big(\langle\theta,\tau\rangle\big)^2 - 4\big(1-\langle\xi,\tau\otimes\tau\rangle\big) \le 0 \qquad \forall\,\tau\in S^{d-1}.$$

Upon dividing by 4 we can then see that the left hand side above may be rewritten as $\big\langle\big(\frac{1}{2}\,\theta\otimes\theta+\xi\big),\tau\otimes\tau\big\rangle$ and thus the condition above is equivalent to $\varrho_+\big(\frac{1}{2}\,\theta\otimes\theta+\xi\big)\le 1$. □

All the ingredients necessary for the proof of inequalities in Theorem 6.9 are already at our disposal:

Proof of Theorem 6.9. We start estimating $\underline{\mathfrak{Z}}$ by using equalities (6.10) and $\underline{Z}_{\underline{F}} = \sup \underline{\mathcal{P}}_{\underline{F}}$, in addition we use the fact that $\widetilde{\mathscr{T}}(\underline{\Omega},f)\supset\mathscr{T}(\underline{\Omega},f)$:

$$\begin{aligned}\underline{\mathfrak{Z}} &= \inf_{\underline{F}\in\mathscr{T}(\underline{\Omega},f)} \underline{Z}_{\underline{F}} \ge \inf_{\underline{F}\in\widetilde{\mathscr{T}}(\underline{\Omega},f)} \underline{Z}_{\underline{F}} \qquad (6.20)\\ &= \inf_{\underline{F}\in\widetilde{\mathscr{T}}(\underline{\Omega},f)}\ \sup_{\underline{v}\in\mathcal{D}(\mathbb{R}^d\setminus\underline{\Upsilon}_0;\mathbb{R}^d)} \left\{\int_{\overline{\underline{\Omega}}}\langle\underline{v},\underline{F}\rangle\ :\ \underline{\varrho}_+\big(\underline{e}(\underline{v})\big)\le 1\ \text{ in }\overline{\underline{\Omega}}\right\}\\ &\ge \sup_{\underline{v}\in\mathcal{D}(\mathbb{R}^d\setminus\underline{\Upsilon}_0;\mathbb{R}^d)}\ \inf_{\underline{F}\in\widetilde{\mathscr{T}}(\underline{\Omega},f)} \left\{\int_{\overline{\underline{\Omega}}}\langle\underline{v},\underline{F}\rangle\ :\ \underline{\varrho}_+\big(\underline{e}(\underline{v})\big)\le 1\ \text{ in }\overline{\underline{\Omega}}\right\} =: \mathcal{Y}.\end{aligned}$$

We may now eliminate the infimum with respect to $\underline{F}$ by employing Lemma 6.10. The function v in the lemma must clearly be an element of $\mathcal{D}(\Omega;\mathbb{R}^{\underline{d}})$ and by decomposing it to $w+\sqrt{2}\,u\,\mathrm{e}_{\underline{d}}$ we infer that

$$\begin{aligned}\mathcal{Y} &= \sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{\int_{\overline{\Omega}}(\sqrt{2}\,u)\,df\ :\ \underline{\varrho}_+\big(\underline{e}(\underline{v})\big)\le 1\ \text{ in }\overline{\underline{\Omega}}\quad\text{for}\quad \underline{v}(\underline{x}) = w(x)+\sqrt{2}\,u(x)\,\mathrm{e}_{\underline{d}}\right\}\\ &= \sqrt{2}\sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}^{d+1})} \left\{\int_{\overline{\Omega}} u\,df\ :\ \varrho\big(\nabla u, e(w)\big)\le 1\ \text{ in }\overline{\underline{\Omega}}\right\} = \sqrt{2}\,\mathcal{Z},\end{aligned}$$

when to pass to the second line we exploited Lemma 6.13 while additionally acknowledging (6.19). The last equality is by definition of ϱ, see Section 5.2.3. The inequality $\underline{\mathfrak{Z}}\ge\sqrt{2}\,\mathcal{Z}$ is thereby established. The second inequality asserted in the theorem follows by equality (6.11). □

The proof above is based on the following equivalence that thanks to Lemma 6.13 holds true for any $(u,w)\in\mathcal{D}(\Omega;\mathbb{R}\times\mathbb{R}^d)$ and $\underline{v}$ satisfying (6.18): the inequality $\underline{\varrho}_+\big(\underline{e}(\underline{v})\big)\le 1$ is satisfied in $\overline{\underline{\Omega}}$ if and only if the inequality $\varrho\big(\nabla u, e(w)\big)\le 1$ is true in $\overline{\Omega}$. This relation will be revisited in Section 6.3.2 for the purposes of analysing the optimally constructed form, although we will be bound to work with Lipschitz continuous functions instead, i.e. with $(u,w)\in\mathrm{Lip}_0(\Omega;\mathbb{R}\times\mathbb{R}^d)$. Then the function $\underline{v}$ defined via (6.18) is clearly an element of $\mathrm{Lip}\big(\underline{\Omega};\mathbb{R}^{\underline{d}}\big)$. By acknowledging Proposition 5.38 it is straightforward to generalize the foregoing equivalence to its non-smooth variant:

$$(u,w)\in\overline{\mathcal{X}}_1 \qquad\Leftrightarrow\qquad \underline{\varrho}_+\big(\underline{e}(\underline{v})\big)\ \le\ 1 \qquad \mathcal{L}^{\underline{d}}\text{-a.e. in }\underline{\Omega} \qquad (6.21)$$

where $\overline{\mathcal{X}}_1$ is defined and characterized in Section 5.3. The result above is insufficient as in Section 6.3.2 we will be working with lower dimensional measures $\underline{\mu} \in \mathcal{M}_+(\overline{\Omega})$. To that aim the $\underline{\mu}$-tangential calculus must be employed:

Lemma 6.14. *For $(u, w) \in \overline{\mathcal{X}}_1 \cap \mathrm{Lip}_0(\Omega; \mathbb{R}^d)$ let us define a function $\underline{v} \in \mathrm{Lip}\big(\underline{\Omega}; \mathbb{R}^{\underline{d}}\big)$ such that*

$$\underline{v}(\underline{x}) := w(x) + \sqrt{2}\, u(x)\, \mathrm{e}_{\underline{d}},$$

i.e. $\underline{v}$ does not depend on the $\underline{d}$-th coordinate. Then, for any measure $\underline{\mu} \in \mathcal{M}_+(\overline{\Omega})$ such that $\underline{\mu}(\Upsilon_0) = 0$ the function $\underline{v}$ is an element of $\mathrm{dom}(\underline{e}_{\underline{\mu}})$ and

$$\underline{\varrho}_{+,\underline{\mu}}\big(\underline{x}, \underline{e}_{\underline{\mu}}(\underline{v})\big) \leq 1 \qquad \underline{\mu}\text{-a.e.}$$

Proof. By construction of $\underline{v}$ and regularity of function (u, w) it is clear that $\underline{v} \in \mathrm{Lip}\big(\Omega; \mathbb{R}^{\underline{d}}\big)$ and $\underline{v} = 0$ on Υ_0. Since $\overline{\Omega}$ is compact the function $\underline{v}$ may be extended to whole $\mathbb{R}^{\underline{d}}$ so that (without relabelling) $\underline{v} \in \mathrm{Lip}_0(\mathbb{R}^d \setminus \Upsilon_0; \mathbb{R}^{\underline{d}})$. From Proposition 2.65 we have $v \in \mathrm{dom}(\nabla_{\underline{\mu}})$ therefore the statement (i) follows from the fact that $\mathrm{dom}(\underline{e}_{\underline{\mu}}) \subset \mathrm{dom}(\nabla_{\underline{\mu}})$, see Proposition 2.66.

To show the assertion (ii) we take a sequence $\big\{(u_n, w_n)\big\} \subset \mathcal{D}(\Omega; \mathbb{R}^{d+1})$ constructed for (u, w) exactly as in the proof of Lemma 5.40, in particular it satisfies conditions (1), (2), (3) therein. We define $\underline{v}_n \in C^\infty\big(\mathbb{R}^{\underline{d}}; \mathbb{R}^{\underline{d}}\big)$ by putting $\underline{v}_n(\underline{x}) = w_n(x) + \sqrt{2}\, u_n(x)\, \mathrm{e}_{\underline{d}}$ and then we can compute that $\underline{e}(\underline{v}_n)(\underline{x}) = e(w_n)(x) + \sqrt{2}\, \nabla u_n(x) \overset{s}{\otimes} \mathrm{e}_{\underline{d}}$. Note that we may redefine $\underline{v}_n$ via multiplying by a cut-off function $\varphi \in \mathcal{D}(\mathbb{R}^{\underline{d}}; \mathbb{R})$ such that $\varphi \equiv 1$ on $\overline{\Omega}$. This way we arrive at $\underline{v}_n \in \mathcal{D}\big(\mathbb{R}^{\underline{d}} \setminus \Upsilon_0; \mathbb{R}^{\underline{d}}\big)$. Below we shall work with such modified functions $\underline{v}_n$. From the property (2) we find that $\sup_n \|\underline{e}(\underline{v}_n)\|_{\infty,\underline{\Omega}} < \infty$ thus, combining with (1), we see that (up to extracting a subsequence) $\underline{v}_n \to \underline{v}$ uniformly in $\overline{\Omega}$ while $\underline{e}(\underline{v}_n) \overset{*}{\rightharpoonup} \underline{\epsilon}$ in $L^\infty_{\underline{\mu}}(\overline{\Omega}; \mathcal{S}^{\underline{d}\times\underline{d}})$. By Lemma 6.13 and the property (3) of the sequence (u_n, w_n) we have $\underline{\varrho}_+\big(\underline{e}(\underline{v}_n)\big) = \varrho\big(\nabla u_n, e(w_n)\big) \leq 1$. The rest of the proof relies on the chain of inequalities as in the proof of Lemma 4.19. □

6.3. Optimal forms constructed from optimal membranes

In this section we will construct in the $\underline{d}$-dimensional domain $\underline{\Omega}$ a form $(z, \underline{\varsigma})$ or $(z, \underline{\lambda})$ that is not only optimal for the form-finding problem (PFF) or (EFF) but also solves the Prager problem (PPP) or (EPP), that is when we take $\underline{F} = \underline{F}_z$. To achieve this goal we will require a complete solution of the corresponding optimal membrane problem posed in d-dimensional region Ω, i.e. we need pairs (u, w) and (ϑ, ς) solving $(\overline{\mathscr{P}})$ and $(\mathscr{P}^*)$, respectively.

The idea for the construction of the form, e.g. $(z, \underline{\varsigma})$ is simple enough. The elevation function z will be taken as the deflection u of the optimal membrane, up to a multiplicative constant. Then the stress $\underline{\varsigma}$ will be computed by "unprojecting" ς onto the thus found surface $\overline{S}_z$. Thanks to the estimates in Theorem 6.9 showing optimality of $(\underline{F}_z, \underline{\varsigma})$ for (PPP) amounts to checking the admissibility and that $\int \mathrm{Tr}\, \underline{\varsigma} \leq \sqrt{2}\, \mathcal{Z}$. Proposing $(\underline{F}_z, \underline{\lambda})$ optimal for (EPP) is then straightforward owing to Proposition 6.5.

Checking the aforementioned admissibility of the pair $(\underline{F}_z, \underline{\varsigma})$, namely checking the equilibrium equation $-\mathrm{Div}\,\underline{\varsigma} = \underline{F}_z$ in $\mathbb{R}^{\underline{d}} \setminus \Upsilon_0$, is the difficult part. It requires certain information on the "differential geometry" of the surface $\overline{S}_z$. The quotation mark is due to the fact that the d-dimensional manifold $\overline{S}_z$ is in general not of C^1 class, meanwhile we need the information $\underline{\mu} := \mathrm{Tr}\,\underline{\varsigma}$ -a.e. where $\underline{\mu}$ can be any Radon measure that is supported on $\overline{S}_z$. At this moment the present author did not succeed in treating the general case, i.e. when (u, w) is an arbitrary pair in $\overline{\mathcal{X}}_1$. Accordingly, Lipschitz continuity of both u and w will be assumed, which renders the manifold $\overline{S}_z$ Lipschitz as well. More importantly, such a regularity assumption unlocks the μ-tangential calculus and optimality conditions in Theorem 5.41, which will be fundamental for proving the equation $-\mathrm{Div}\,\underline{\varsigma} = \underline{F}_z$.

6.3.1. The construction of the form

Let us assume that the bounded domain $\Omega \subset \mathbb{R}^d$ is star-shaped with respect to a ball (in order to exploit results from Chapter 5), while f is any element of $\mathcal{M}(\overline{\Omega}; \mathbb{R})$. We shall also make an assumption on the "height" of the $\underline{d}$-dimensional domain $\underline{\Omega} = \Omega \times (-h_0, h_0)$, namely

$$h_0 \geq \frac{1}{2}\,\mathrm{diam}(\Omega) \tag{6.22}$$

where $\mathrm{diam}(\Omega)$ stands for the diameter of the set Ω. Throughout the rest of Section 6.3 we choose and fix two pairs

$$(u, w) \in \overline{\mathcal{X}}_1 \quad \text{and} \quad (\vartheta, \varsigma) \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d \times \mathcal{S}^{d\times d}_+) \qquad \text{that solve} \qquad (\overline{\mathscr{P}}) \ \text{ and } \ (\mathscr{P}^*),$$

respectively, where the pair of problems $(\overline{\mathscr{P}}), (\mathscr{P}^*)$ are posed in the CanMD setting for data Ω, f. Note that solutions as above always exist owing to Corollary 5.26.

The construction of an optimal form begins with the elevation function:

$$z := \frac{1}{\sqrt{2}}\,u, \qquad \underline{z}(x) := \big(x, z(x)\big) = x + z(x)\,\mathrm{e}_{\underline{d}}. \tag{6.23}$$

Obviously $z \in C_0(\Omega)$ due to the properties of the set $\overline{\mathcal{X}}_1$, see Theorem 5.32. We can see that $\underline{z} : \overline{\Omega} \to \overline{S}_z$ is a homeomorphism (although z is not), in particular $\underline{z}^{-1} : \overline{S}_z \to \overline{\Omega}$ is well defined and continuous, in fact: $\underline{z}^{-1} = \pi_{\mathbb{R}^d}$. The assumption (6.22) on h_0 becomes clear: since $\|u\|_{\infty,\Omega} \leq \frac{1}{\sqrt{2}}\,\mathrm{diam}(\Omega)$ by assertion (i) in Lemma 5.24 we readily check that

$$\overline{S}_z = \Big\{ \big(x, z(x)\big) \,:\, x \in \overline{\Omega} \Big\} \quad \subset \quad \overline{\underline{\Omega}} = \overline{\Omega} \times [-h_0, h_0].$$

Next we propose a particular representation $(\vartheta, \varsigma) = (q\mu, \sigma\mu)$ where $(q, \sigma) \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d \times \mathcal{S}^{d\times d}_+)$:

$$\mu := \varrho^0(\vartheta, \varsigma) \ \in \ \mathcal{M}_+(\overline{\Omega}), \qquad q := \frac{d\vartheta}{d\mu}, \qquad \sigma := \frac{d\varsigma}{d\mu} \tag{6.24}$$

where ϱ^0 is the convex, one-homogeneous and lower semi-continuous support function for the set C (see Section 5.2.3); for $(q, \sigma) \in \mathbb{R}^d \times \mathcal{S}^{d\times d}_+$ we have $\varrho^0(q, \sigma) = \mathrm{Tr}\,\sigma + \frac{1}{2}\langle \sigma^{-1} q, q\rangle$. Clearly $\mu(\overline{\Omega}) = \int_{\overline{\Omega}} \varrho^0(\vartheta, \varsigma) = \mathcal{Z} < \infty$, therefore $\mu \in \mathcal{M}_+(\overline{\Omega})$ as stated above.

We may now construct the $\underline{d}$-dimensional measures and functions, starting with:

$$\underline{\mu} := \sqrt{2}\,\underline{z}\,\#\,\mu, \qquad \underline{\sigma} := \begin{bmatrix} \sigma & \frac{1}{\sqrt{2}}\,q \\ \frac{1}{\sqrt{2}}\,q^\top & \frac{1}{2}\,\langle \sigma^{-1}q, q\rangle \end{bmatrix} \circ \underline{z}^{-1}. \tag{6.25}$$

Below we show that $\underline{\mu}$ and $\underline{\sigma}$ are well-defined and prove some properties that will be useful in further steps:

Proposition 6.15. *The following statements hold true:*

(i) $\underline{\mu}$ *is an element of* $\mathcal{M}_+(\overline{S}_z) \subset \mathcal{M}_+(\overline{\Omega})$ *while* $\underline{\sigma} : \overline{S}_z \to \mathcal{S}^{\underline{d}\times\underline{d}}$ *is a* $\underline{\mu}$*-measurable function;*

(ii) Let $\hat{\theta} : \overline{\Omega} \to \mathbb{R}^d$ *be the* μ*-measurable function as in Proposition 5.13. Then*

$$\underline{\sigma} = \left(\begin{bmatrix} \mathrm{I} & \frac{1}{\sqrt{2}}\,\hat{\theta} \end{bmatrix}^\top \sigma \begin{bmatrix} \mathrm{I} & \frac{1}{\sqrt{2}}\,\hat{\theta} \end{bmatrix} \right) \circ \underline{z}^{-1}. \tag{6.26}$$

As a result for $\underline{\mu}$*-a.e.* $\underline{x}$ *we have* $\underline{\sigma}(\underline{x}) \in \mathcal{S}^{\underline{d}\times\underline{d}}_+$ *while*

$$\operatorname{Im}\underline{\sigma}(\underline{x}) \;=\; \mathfrak{T}(\underline{x}) \;:=\; \left\{ \tau + \tfrac{1}{\sqrt{2}}\langle \hat{\theta}(x), \tau\rangle\, \mathrm{e}_{\underline{d}} \,:\, \tau \in \operatorname{Im}\sigma(x) \right\} \tag{6.27}$$

where $x = \underline{z}^{-1}(\underline{x}) = \pi_{\mathbb{R}^d}(\underline{x})$*. In particular,* $\operatorname{rank}\big(\underline{\sigma}(\underline{x})\big) \le \operatorname{rank}\big(\sigma(x)\big) \le d$*.*

(iii) The stress function $\underline{\sigma}$ *is of unit trace, i.e.*

$$\operatorname{Tr}\underline{\sigma} = 1 \qquad \underline{\mu}\text{-a.e.} \tag{6.28}$$

hence $\underline{\sigma} \in L^\infty_{\underline{\mu}}(\overline{S}_z; \mathcal{S}^{\underline{d}\times\underline{d}}_+)$*.*

Proof. Since $\underline{z} : \overline{\Omega} \to \overline{S}_z$ is continuous the push-forward measure $\underline{\mu}$ is a positive Radon (see Section 2.2), i.e. $\underline{\mu} \in \mathcal{M}_+(\overline{S}_z)$. To proof the second claim of (i) let us put $g = \underline{z}^{-1}$ being a continuous function from $\overline{S}_z$ (endowed with the relative topology) to $\overline{\Omega}$. Let us denote the block matrix in definition (6.25) by M; then $M : \overline{\Omega} \to \mathcal{S}^{\underline{d}\times\underline{d}}$ is a μ-measurable function (since all the components of the block matrix are), i.e. for any open set $\mathscr{V} \subset \mathcal{S}^{\underline{d}\times\underline{d}}$ the pre-image $M^{-1}(\mathscr{V})$ is μ-measurable therefore it is equal to $B \cup N \subset \overline{\Omega}$ where B is Borel and N is μ-negligible. Then, since $\underline{\sigma} = M \circ g$ we have $\underline{\sigma}^{-1}(\mathscr{V}) = g^{-1}\big(M^{-1}(\mathscr{V})\big) = g^{-1}\big(B \cup N\big) = g^{-1}(B) \cup g^{-1}(N)$. Since g is continuous the set $g^{-1}(B)$ is a Borel subset of $\overline{S}_z$, while, by definition of the push-forward, $\underline{\mu}\big(g^{-1}(N)\big) = \underline{\mu}\big(\underline{z}(N)\big) = \mu(N) = 0$. We infer that $\underline{\sigma}^{-1}(\mathscr{V})$ is $\underline{\mu}$-measurable thus concluding the proof of assertion (i).

Showing (6.26) reduces to linear algebra and acknowledging that $q = \sigma\hat{\theta}$ and $\frac{1}{2}\,\langle \sigma^{-1}q, q\rangle = \frac{1}{2}\,\langle \sigma\hat{\theta}, \hat{\theta}\rangle$, cf. Proposition 5.13. Then positive semi-definiteness of $\underline{\sigma} = (A^\top \sigma A) \circ \underline{z}^{-1}$ follows from positive semi-definiteness of σ, while (6.27) is straightforward, hence assertion (ii) is established.

To show the claim (iii) we may work with the block matrix in (6.25) directly: clearly μ-a.e. $\operatorname{Tr}\underline{\sigma}\big(\underline{z}(\,\cdot\,)\big) = \operatorname{Tr}\sigma + \frac{1}{2}\,\langle \sigma^{-1}q, q\rangle = \underline{\varrho}^0(q, \sigma) = 1$, while the last equality exploits the fact that $\mu = \underline{\varrho}^0(q\mu, \sigma\mu) = \underline{\varrho}^0(q, \sigma)\,\mu$, see (6.24). □

The pair $(\underline{\varsigma}, \underline{F})$ being a candidate for solving the plastic setting of Prager problem (PPP) may be readily proposed:

$$\underline{\varsigma} := \underline{\sigma}\underline{\mu}, \qquad \underline{F} := \underline{F}_z = \mathrm{e}_{\underline{d}}\,\underline{z}\,\#\,f. \tag{6.29}$$

Should the pair above solve (PPP), Proposition 6.5 immediately furnishes a recipe for a pair $(\underline{\lambda}, \underline{F})$ that solves (EPP): we must choose a Hooke function $\underline{\mathscr{H}} \in L^\infty_{\underline{\mu}}(\overline{S}_z; \mathcal{K}_{fib})$ such that for $\underline{\mu}$-a.e. $\underline{x} \in \overline{S}_z$ there holds $\underline{\mathscr{H}}(\underline{x}) \in \underline{\mathcal{K}}_{1,+}\big(\underline{\sigma}(\underline{x})\big)$. Since we are within the CanMD anisotropy setting, after Example 3.28 this condition may be rewritten as

$$\underline{\mathscr{H}}(\underline{x}) \in \left\{ \sum_{i=1}^m \alpha_i \, \underline{\tau}_i \otimes \underline{\tau}_i \otimes \underline{\tau}_i \otimes \underline{\tau}_i \ : \ \underline{\tau}_i \in S^{d-1} \cap \operatorname{Im} \underline{\sigma}(\underline{x}), \ \sum_{i=1}^m \alpha_i = 1, \, \underline{\sigma}(\underline{x}) = \sum_{i=1}^m \alpha_i \, \underline{\tau}_i \otimes \underline{\tau}_i \right\} \tag{6.30}$$

where we used the fact that $\underline{\mu}$-a.e. $\operatorname{Tr}\big(\underline{\sigma}(\underline{x})\big) = 1$, see the assertion (iii) of Proposition 6.15. Note that such function $\underline{\mathscr{H}}$ may be always chosen so that it is $\underline{\mu}$-measurable thanks to Lemma 3.41. Since $\operatorname{Tr} \underline{\varsigma} = \underline{\mu}$ we can follow Proposition 6.5 to propose a candidate pair $(\underline{\lambda}, \underline{F})$ for solving the elastic setting of the Prager problem (EPP):

$$\underline{\lambda} := \frac{C_0}{\sqrt{2}\, \mathcal{Z}} \, \underline{\mathscr{H}} \underline{\mu}, \qquad \underline{F} := \underline{F}_z = \mathrm{e}_{\underline{d}} \, \underline{z} \, \# \, f;$$

the value $\sqrt{2}\,\mathcal{Z}$ in the denominator is the predicted value of $\underline{\mathfrak{Z}}$, i.e. we expect that the inequalities in Theorem 6.9 are in fact equalities.

Remark 6.16. According to assertion (ii) of Proposition 6.15 in the set of optimal Hooke tensors (6.30) the space $\operatorname{Im} \underline{\sigma}(\underline{x})$ may be $\underline{\mu}$-a.e. replaced by $\mathfrak{T}(\underline{x})$ defined in (6.27). Consequently, the vectors $\underline{\tau}_i$ used for constructing the tensor $\underline{\mathscr{H}}(\underline{x})$ must be always chosen from $\mathfrak{T}(\underline{x})$ being a proper subspace of $\mathbb{R}^d$ that is at most d-dimensional. Loosely speaking, we extract an additional information on the dimensionality of tensor $\underline{\mathscr{H}}(\underline{x})$ and on the orientation of the fibres that make it up.

The construction of the forms is complete but this is as far as we can go without making additional assumptions. At the moment the optimality of $(\underline{\varsigma}, \underline{F}_z)$ – or rather the admissibility for (PPP) – cannot be proved unless we are in control of the μ-tangential gradient of the elevation function z. To this aim we make

$$\text{an assumption:} \quad \text{the chosen solution } (u, w) \text{ of } (\overline{\mathscr{P}}) \text{ is Lipschitz continuous.} \tag{6.31}$$

We stress that although expecting $(u, w) \in \operatorname{Lip}_0(\Omega; \mathbb{R} \times \mathbb{R}^d)$ is not based on any analysis, such additional regularity of (u, w) so far proved to be the case in all examples (analytical or numerical) where Ω was convex, see Section 5.7.2 for more details.

With the solution (u, w) being an element of $\operatorname{Lip}_0(\Omega; \mathbb{R} \times \mathbb{R}^d)$ we may now exploit the results of Section 5.4. In particular we know that $u \in \operatorname{dom}(\nabla_\mu)$ and $w \in \operatorname{dom}\big((\nabla_\mu)^d\big) \subset \operatorname{dom}\big(e_\mu\big)$ intended for the domain $\mathscr{U} = \Omega$ (see Section 2.5). Automatically we infer that $z = \frac{1}{\sqrt{2}}\, u$ is an element of $\operatorname{dom}(\nabla_\mu)$ and that $\nabla_\mu z = \frac{1}{\sqrt{2}} \nabla_\mu u$. Finally, we can make use of Theorem 5.41: since the pairs $(u, w) \in \operatorname{Lip}_0(\Omega; \mathbb{R} \times \mathbb{R}^d)$ and (ϑ, ς) are solutions of $(\overline{\mathscr{P}})$ and, respectively, $(\mathscr{P}^*)$ we know that they must satisfy the system of optimality conditions (5.67). In particular, we extract the equation (v) in (5.67), namely $q = \sigma \nabla_\mu u$, which provides the missing piece of information:

$$\frac{1}{\sqrt{2}}\, q = \sigma \nabla_\mu z \qquad \text{by the optimality condition (v) in (5.67).} \tag{6.32}$$

Regularity of the function $z : \overline{\Omega} \to \mathbb{R}$ translates to the function $\underline{z} = (\mathrm{id}, z) : \overline{\Omega} \to \mathbb{R}^{\underline{d}}$, namely

$$\underline{z} \in \mathrm{Lip}(\Omega; \mathbb{R}^{\underline{d}}) \ \subset \ \mathrm{dom}\big((\nabla_\mu)^{\underline{d}}\big), \qquad \nabla_\mu \underline{z} = \Big[P_\mu \quad \nabla_\mu z \Big]. \tag{6.33}$$

Note that $\underline{z}$ does not vanish on $\partial\Omega$ therefore the operator ∇_μ is applied to $\underline{z}$ for the domain $\mathscr{U} = \mathbb{R}^d$. The block-matrix representation of $\nabla_\mu \underline{z}$ is due to the fact that ∇_μ is applied column-wise (see Remark 2.63) while $\nabla_\mu(\mathrm{id}) = P_\mu\big(\nabla(\mathrm{id})\big) = P_\mu \, \mathrm{I} = P_\mu$. Note that in the block matrix $\nabla_\mu z$ is *a priori* computed for the μ-tangential calculus in the domain $\mathscr{U} = \mathbb{R}^d$ although, thanks to Remark 2.53 and to the fact that $\mu(\mathbb{R}^d \setminus \Omega) = 0$ by condition (iii) in (5.67), such function $\nabla_\mu z$ coincides μ-a.e. with the one computed for the domain $\mathscr{U} = \Omega$ as it is intended in (6.32). As the next step we observe that μ-a.e.

$$\big(\nabla_\mu \underline{z}\big)^\top \sigma \ \nabla_\mu \underline{z} = \begin{bmatrix} P_\mu \sigma P_\mu & P_\mu(\sigma \nabla_\mu z) \\ \big(P_\mu(\sigma \nabla_\mu z)\big)^\top & \langle \sigma \nabla_\mu z, \nabla_\mu z \rangle \end{bmatrix} = \underline{\varsigma}\big(\underline{z}(\,\cdot\,)\big). \tag{6.34}$$

Considering the block form of $\nabla_\mu \underline{z}$ the first equality is purely algebraic (recall that P_μ is a point-wise symmetric matrix) and we see that the relation above lies closely to (6.26). This is partially clear since due to optimality condition (v) in (5.67) the function $\hat{\theta}$ may be always chosen as $\nabla_\mu u$ and then $\frac{1}{\sqrt{2}} \hat{\theta} = \nabla_\mu z$. In addition we must justify that I may be replaced by P_μ. We recall that since $\varsigma = \sigma\mu$ satisfies the equilibrium $\mathrm{Div}\, \varsigma = 0$ in Ω we know that $\sigma \in M^e_\mu$ μ-a.e., see Theorem 2.57. In turn, by Proposition 2.66 we know that $P_\mu \sigma P_\mu = \sigma$ and moreover that $\mathrm{Im}\, \sigma \subset T_\mu$ μ-a.e. As a result we may delete P_μ in the quadratic block matrix above.

With the formula (6.34) at our disposal we may revisit the claim (ii) in Proposition 6.15 in order to relate the image of $\underline{\varsigma}$ to the geometry of the surface S_z. For $\underline{\mu}$-a.e. $\underline{x} \in S_z$ we have

$$\mathrm{Im}\, \underline{\varsigma}(\underline{x}) \ = \ \mathfrak{T}(\underline{x}) \ = \ \Big\{ \tau + \big\langle \nabla_\mu z(x), \tau \big\rangle \, \mathrm{e}_{\underline{d}} \ : \ \tau \in \mathrm{Im}\, \sigma(x) \Big\} \ \subset \ \mathrm{Im}\, \big(\nabla_\mu \underline{z}(x)\big)^\top$$

where $x = \underline{z}^{-1}(\underline{x})$. Finally, if z is C^1 then $\big(\nabla_\mu \underline{z}(x)\big)^\top = \big(\nabla \underline{z}(x)\big)^\top P_\mu$ and we find that

$$z \in C^1(\overline{\Omega}) \qquad \Rightarrow \qquad \mathrm{Im}\, \underline{\varsigma}(\underline{x}) \ \subset \ \mathrm{Im}\, \big(\nabla \underline{z}(x)\big)^\top \ = \ T_{S_z}(\underline{x}) \tag{6.35}$$

where $T_{S_z}(\underline{x})$ is the classically understood tangent space at $\underline{x}$ to S_z. Similar conclusions may be drawn about the optimal Hooke tensor function $\mathscr{H}$, cf. Remark 6.16.

We are ready to show that the pair $(\underline{\varsigma}, \underline{F}_z)$ is admissible for the Prager problem (PPP):

Lemma 6.17. *The stress field $\underline{\varsigma}$ and the load $\underline{F}_z$ satisfy the equilibrium equation:*

$$-\mathrm{Div}\, \underline{\varsigma} = \underline{F}_z \qquad \textit{in} \quad \mathcal{D}'(\mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0; \mathbb{R}^{\underline{d}}).$$

Proof. We must prove that for all test functions $\underline{\psi} \in \mathcal{D}(\mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0; \mathbb{R}^{\underline{d}})$ there holds $\int_{\underline{\Omega}} \big\langle \underline{\nabla} \underline{\psi}, \underline{\varsigma} \big\rangle = \int_{\underline{\Omega}} \big\langle \underline{\psi}, \underline{F}_z \big\rangle$. Let us put $\psi = \underline{\psi} \circ \underline{z}$ being a function in $\mathrm{Lip}_0(\Omega; \mathbb{R}^{\underline{d}})$ which is an element of $\mathrm{dom}\big((\nabla_\mu)^{\underline{d}}\big)$ and by virtue of Proposition 2.64 (set $u := \underline{\psi}$, $G := \underline{z}$, $v := \psi$) the μ-tangential chain rule follows:

$$\nabla_\mu \psi(x) = \nabla_\mu \underline{z}(x) \, \underline{\nabla} \underline{\psi}\big(\underline{z}(x)\big) \qquad \text{for } \mu\text{-a.e. } x.$$

We can make a decomposition $\psi = \phi + \varphi\, \mathrm{e}_{\underline{d}}$ where $\phi \in \mathrm{Lip}_0(\Omega;\mathbb{R}^d)$ and $\varphi \in \mathrm{Lip}_0(\Omega;\mathbb{R})$; then, in the block format, $\nabla_\mu \psi = \begin{bmatrix}\nabla_\mu \phi & \nabla_\mu \varphi\end{bmatrix}$. Recalling the relation between $\underline{\sigma}$ and σ in (6.34), using the equality $\big\langle A, B^\top C B\big\rangle = \langle BA, CB\rangle$ and acknowledging the block form in (6.33) we discover that μ-a.e.

$$\begin{aligned}\Big\langle \underline{\nabla}\underline{\psi}\big(\underline{z}(\,\cdot\,)\big)\,,\, \underline{\sigma}\big(\underline{z}(\,\cdot\,)\big)\Big\rangle = \Big\langle \nabla_\mu\psi\,,\, \begin{bmatrix}\sigma P_\mu & \sigma\,\nabla_\mu z\end{bmatrix}\Big\rangle &= \Big(\langle \nabla_\mu\phi, \sigma\rangle + \langle \nabla_\mu\varphi, \sigma\,\nabla_\mu z\rangle\Big) \\ &= \Big(\langle e_\mu(\phi), \sigma\rangle + \tfrac{1}{\sqrt{2}}\langle \nabla_\mu\varphi, q\rangle\Big)\end{aligned} \tag{6.36}$$

where to pass to the last line we again used the optimality condition (6.32) while by virtue of Proposition 2.66 we have $\phi \in \mathrm{dom}(e_\mu)$ and $\langle \nabla_\mu\phi, \sigma\rangle = \big\langle \frac{1}{2}\big(\nabla_\mu\phi + (\nabla_\mu\phi)^\top\big), \Pi^e_\mu\sigma\big\rangle = \langle e_\mu(\phi), \sigma\rangle$. By the change of variable formula in the general, push-forward setting (2.6)

$$\begin{aligned}&\int_{\underline{\Omega}} \big\langle \underline{\nabla}\underline{\psi}, \underline{\varsigma}\big\rangle = \int_{\underline{\Omega}} \big\langle \underline{\nabla}\underline{\psi}, \underline{\sigma}\big\rangle\, d\underline{\mu} = \sqrt{2}\int_{\overline{\Omega}} \Big\langle \underline{\nabla}\underline{\psi}\big(\underline{z}(\,\cdot\,)\big)\,,\, \underline{\sigma}\big(\underline{z}(\,\cdot\,)\big)\Big\rangle\, d\mu \\ &= \int_{\overline{\Omega}} \Big(\langle e_\mu(\phi), \sqrt{2}\,\sigma\rangle + \langle \nabla_\mu\varphi, q\rangle\Big)\, d\mu = \int_{\overline{\Omega}} \varphi\, df = \int_{\underline{\Omega}} \big\langle \underline{\psi}(\underline{x}), \mathrm{e}_{\underline{d}}\big\rangle\, (\underline{z}\#f)(d\underline{x}) = \int_{\underline{\Omega}} \big\langle \underline{\psi}, \underline{F}_z\big\rangle.\end{aligned}$$

Above we have exploited the μ-tangential integration by part formula given in (5.63): it holds true here since $-\mathrm{div}(q\mu) = f$ and $\mathrm{Div}\,(\sqrt{2}\,\sigma\mu) = \sqrt{2}\,\mathrm{Div}(\sigma\mu) = 0$ in Ω. The last equality is by construction of $\underline{F}_z$. □

The main theorem of this chapter is now a simple consequence of the hitherto obtained results:

Theorem 6.18 (**Construction of a form solving the Prager problem**). *Assume that the solution (u, w) of the problem $(\overline{\mathscr{P}})$ is Lipschitz continuous and that $h_0 \geq \frac{1}{2}\,\mathrm{diam}(\Omega)$.*

Then the pair

$$\underline{\varsigma} = \underline{\sigma}\underline{\mu}, \qquad \underline{F} = \mathrm{e}_{\underline{d}}\, \underline{z}\,\#\, f \qquad \text{that is supported on } \overline{S}_z$$

solves the plastic setting of the Prager problem (PPP) *with $\underline{\mathfrak{Z}} = \sqrt{2}\,\mathcal{Z}$, while the pair*

$$\underline{\lambda} = \frac{C_0}{\sqrt{2}\,\mathcal{Z}}\, \mathscr{H}\underline{\mu}, \qquad \underline{F} = \mathrm{e}_{\underline{d}}\, \underline{z}\,\#\, f \qquad \text{that is supported on } \overline{S}_z$$

solves the elastic setting of the Prager problem (EPP) *with $\mathfrak{C} = \frac{\mathcal{Z}^2}{C_0}$.*

Proof. By Lemma 6.17 the pair $(\underline{\varsigma}, \underline{F})$ is admissible for (PPP). Thanks to Theorem 6.9 to prove optimality of this pair it is enough to show that $\int_{\underline{\Omega}} \mathrm{Tr}\,\underline{\varsigma} = \sqrt{2}\,\mathcal{Z}$. Considering assertion (iii) in Proposition 6.15 we may easily compute

$$\int_{\underline{\Omega}} \mathrm{Tr}\,\underline{\varsigma} = \int_{\underline{\Omega}} \mathrm{Tr}\,\underline{\sigma}\, d\underline{\mu} = \int_{\underline{\Omega}} d\underline{\mu} = \sqrt{2}\int_{\overline{\Omega}} d\mu = \sqrt{2}\int_{\overline{\Omega}} \varrho^0(\vartheta, \varsigma) = \sqrt{2}\,\mathcal{Z}$$

where the last equality acknowledges the optimality of (ϑ, ς) for $(\mathscr{P}^*)$. Since eventually $\underline{\mathfrak{Z}} = \sqrt{2}\,\mathcal{Z}$ the optimality of the pair $(\underline{\lambda}, \underline{F})$ for (EPP) follows by Proposition 6.5. □

Owing to the $c_{\mathbf{v}}$-optimal transport theory that underlies the optimal membrane problem (cf. Section 5.6) we can now deduce that for a load f of one sign the optimal form lies on one side of the base plane $\mathbb{R}^d$:

Corollary 6.19. *Assume that $f \geq 0$ (resp. $f \leq 0$) and that the prerequisites of Theorem 6.18 are met. Let us put $\mathbf{v} = \mathrm{id} - \mathbf{i}^{-1}(w) \in \mathbf{V}_\Omega$. Then:*

$$z(x) = \frac{1}{\sqrt{2}}\, d^{\partial\Omega}_{c_{\mathbf{v}}}(x) \geq 0 \qquad \Big(\textit{resp.} \quad z(x) = -\frac{1}{\sqrt{2}}\, d^{\partial\Omega}_{c_{\mathbf{v}}}(x) \leq 0\Big) \qquad \forall\, x \in \operatorname{spt} f.$$

where $d^{\partial\Omega}_{c_{\mathbf{v}}}(x) = \inf_{y\in\partial\Omega} c_{\mathbf{v}}(x,y)$.

Moreover, there always exists a pair $(\underline{\varsigma}, \underline{F})$ solving the Prager problem (PPP) *that satisfies*

$$\operatorname{spt}\underline{\varsigma} \subset \overline{\Omega}\times[0,h_0] \qquad \Big(\textit{resp.} \quad \operatorname{spt}\underline{\varsigma} \subset \overline{\Omega}\times[-h_0,0]\Big).$$

The analogous property holds for solution $(\underline{\lambda}, \underline{F})$ of (EPP).

Proof. Due to the relation $z = \frac{1}{\sqrt{2}}\, u$ the first assertion is a simple consequence of Proposition 5.84 and Corollary 5.89. The "moreover part" is achieved by redefining u: we put $\tilde{u} = d^{\partial\Omega}_{c_{\mathbf{v}}}$. Then, by virtue of Proposition 5.87 the pair $(\tilde{u}, w) \in \overline{\mathcal{X}}_1$ solves $(\overline{\mathscr{P}})$ as well while $\tilde{u} \geq 0$. In addition $\tilde{u} \in \mathrm{Lip}_0(\Omega;\mathbb{R})$ thanks to $w \in \mathrm{Lip}_0(\Omega;\mathbb{R}^d)$ and characterization (iii) of the set $\overline{\mathcal{X}}_1$ in Theorem 5.32. We may thus conduct the construction folr $\tilde{z} = \frac{1}{\sqrt{2}}\,\tilde{u}$ and the claim follows. □

Considering the inclusions (6.15), (6.16) it is straightforward that the obtained structures solve the optimal form-finding problems that were a point of departure in this chapter, see Section 6.1:

Corollary 6.20. *Under the prerequisites of Theorem 6.18 the pairs $(z, \underline{\varsigma})$ and $(z, \underline{\lambda})$ solve the optimal form-finding problems: the plastic setting* (PFF) *and, respectively, the elastic setting* (EFF). *In addition* $\min(\mathrm{PFF}) = \underline{\mathfrak{Z}} = \sqrt{2}\,\mathcal{Z}$ *and* $\min(\mathrm{EFF}) = \underline{\mathfrak{C}} = \frac{\mathcal{Z}^2}{C_0}$.

In the course of proving Theorem 6.18 we find a more general result that is rather unexpected. In both problems (PPP) and (EPP) the designed structure is allowed to rest only on the plane curve $\underline{\Upsilon}_0 = \partial\Omega \times \{0\}$. In the proof of inequality $\underline{\mathfrak{Z}} \geq \sqrt{2}\,\mathcal{Z}$ in Theorem 6.9 (cf. the chain (6.20)) this is reflected in taking the supremum among all functions $\underline{v} \in \mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0$. Eventually the number $\mathcal{Y}$ at the end of the chain is recognized as the supremum with respect to functions $\underline{v}$ that in $\overline{\Omega}$ do not depend on the $\underline{d}$-th variable at all and, as a result, they must be zero on the whole cylinder $\partial\Omega\times[-h_0,h_0]$. We infer that this stronger Dirichlet condition could be imposed right from the start, in other words $\sqrt{2}\,\mathcal{Z}$ would remain a lower bound for $\underline{\mathfrak{Z}}$ even if we allowed the structure to rest on this cylinder. The idea behind the proof of Theorem 6.18 was to give a recipe for an admissible structure that saturates the lower bound $\sqrt{2}\,\mathcal{Z}$, which works in this broader design problem. Ultimately we conclude that:

Corollary 6.21. *Under the prerequisites of Theorem 6.18 the pair $(\underline{\varsigma}, \underline{F})$ solves the variant of the problem* (PPP) *where the structure may rest on the whole cylinder $\partial\Omega\times[-h_0,h_0]$, i.e. it solves the problem*

$$\min_{\substack{\underline{F}\in\mathcal{M}(\overline{\Omega};\mathbb{R}^{\underline{d}})\\ \underline{\varsigma}\in\mathcal{M}(\overline{\Omega};\mathcal{S}^{\underline{d}\times\underline{d}}_+)}} \left\{ \int \mathrm{Tr}\,\underline{\varsigma} \;:\; \underline{F} \in \mathscr{T}(\Omega,f), \quad -\mathrm{Div}\,\underline{\varsigma} = \underline{F} \quad \textit{in } \mathbb{R}^{\underline{d}}\setminus\big(\partial\Omega\times[-h_0,h_0]\big) \right\} \tag{6.37}$$

An analogous statement holds true for the problem (EPP).

This corollary settles yet another design problem where together with the structure we seek an optimal kinematical support being a $\underline{d}$-dimensional curve $\widetilde{\Upsilon}_0$ obtained by elevating the plane curve $\partial\Omega \times \{0\}$. Corollary 6.21 shows that with all the cylinder $\partial\Omega \times [-h_0, h_0]$ available the choice $\widetilde{\Upsilon}_0 = \Upsilon_0 = \partial\Omega \times \{0\}$ is still optimal. The question of finding the optimal supporting curve $\widetilde{\Upsilon}_0$ is therefore trivial. A completely different design problem would be to search for the optimal structure with a fixed, prescribed supporting curve $\widetilde{\Upsilon}_0$ – to this topic we will devote a separate Section 6.6.3.4.

6.3.2. Analysis of the optimal elastic form – computing the deformation

In the process of proving the key estimates from Theorem 6.9 an important role was played by the $\underline{d}$-dimensional displacement fields that do not vary in the $\underline{d}$-th coordinate. We shall now show that if for functions (u, w) solving $(\overline{\mathscr{P}})$ we define the function $\underline{v} : \overline{\Omega} \to \mathbb{R}^{\underline{d}}$ as below

$$\underline{v}(\underline{x}) := w(x) + \sqrt{2}\, u(x)\, \mathrm{e}_{\underline{d}} \tag{6.38}$$

then, up to a multiplicative constant, $\underline{v}$ is the deformation of the optimal elastic form $\underline{\lambda}$ subject to the optimally transmitted load $\underline{F}_z$. Let us stress we will stick to the assumption (6.31) on Lipschitz regularity of (u, w) and all the objects $u, w, \vartheta, \varsigma, \mu, \sigma, q, z, \underline{\mu}, \underline{\sigma}, \underline{\varsigma}, \mathscr{H}, \underline{\lambda}, \underline{F}_z$ are chosen precisely as in the previous section.

Since by Theorem 6.18 the pair $(\underline{\varsigma}, \underline{F}_z)$ solves the Prager problem (PPP), thanks to Proposition 6.4 we know that $\underline{\varsigma}$ solves the problem $(\underline{\mathcal{P}}^*_{\underline{F}}) = (\underline{\mathcal{P}}^*_{\underline{F}_z})$. Similarly we find that $\underline{\lambda}$ solves the Free Material Design problem (6.6) for $\underline{F} = \underline{F}_z$. Of course the function $\underline{v}$ – being an element of the space $\mathrm{Lip}(\underline{\Omega}; \mathbb{R}^{\underline{d}})$ – cannot in general solve the problem $(\underline{\mathcal{P}}_{\underline{F}})$ where only smooth functions $\underline{v} \in \underline{\mathcal{U}}_1$ are considered. This is why we relax the latter problem in the spirit of Section 4.3: we define $\overline{\underline{\mathcal{U}}}_1$ as the closure of $\underline{\mathcal{U}}_1$ (see (6.12)) with respect to the topology of uniform convergence on $\overline{\underline{\Omega}}$ and then we formulate

$$\sup\left\{\int_{\underline{\Omega}} \langle \underline{v}, \underline{F}_z\rangle \;:\; \underline{v} \in \overline{\underline{\mathcal{U}}}_1\right\} \tag{$\overline{\underline{\mathcal{P}}}_{\underline{F}_z}$}$$

The integral functional being maximized is continuous with respect to uniform convergence, therefore

$$\sup \overline{\underline{\mathcal{P}}}_{\underline{F}_z} = \sup \underline{\mathcal{P}}_{\underline{F}_z} = \min \underline{\mathcal{P}}^*_{\underline{F}_z} = \underline{Z}_{\underline{F}_z} = \mathfrak{Z} = \sqrt{2}\, \mathcal{Z}$$

and thus we deduce that:

Proposition 6.22. *The displacement function $\underline{v}$ in* (6.38) *and the stress field $\underline{\varsigma}$ solve the problems $(\overline{\underline{\mathcal{P}}}_{\underline{F}_z})$ and $(\underline{\mathcal{P}}^*_{\underline{F}_z})$ respectively.*

Proof. The fact that $\underline{v} \in \overline{\underline{\mathcal{U}}}_1$ is due to the construction of the sequence $\{\underline{v}_n\} \subset \underline{\mathcal{U}}_1$ in the proof of Lemma 6.14. Then, by the change of variable formula we have $\int_{\overline{\Omega}} \langle \underline{v}, \underline{F}_z\rangle = \int_{\overline{\Omega}} \langle \underline{v}\big(\underline{z}(x)\big), \mathrm{e}_{\underline{d}}\rangle\, f(dx) = \int_{\overline{\Omega}} \sqrt{2}\, u(x)\, f(dx) = \sqrt{2}\, \mathcal{Z}$ where the latter equality acknowledges that (u, w) solves $(\overline{\mathscr{P}})$. □

With $\underline{v}$ being a solution of $(\overline{\underline{\mathcal{P}}}_{\underline{F}_z})$ one would like to make use of the tools developed in Sections 4.3 and 4.4. For instance, to exploit Proposition 4.13 in order to show that the

function $\underline{\breve{v}} = \frac{Z_{\underline{F}_z}}{C_0}\,\underline{v}$ solves the relaxed maximization problem defining compliance $\underline{\mathcal{C}}(\underline{\lambda}, \underline{F}_z)$ in (6.2) or to use Corollary 4.24 in which $\underline{\breve{v}}$ satisfies the system of PDEs of elasticity. Employing those results directly would be unjustified since, in short, the set $\overline{\underline{\mathcal{U}}}_1$ is not compact in $C(\overline{\underline{\Omega}}; \underline{\mathbb{R}}^d)$ – which is the prerequisite in Section 4.3 – and the function j_+ does not meet the ellipticity condition ($\mathrm{A}_{\mathrm{ellip}}$), thus ruling out the usage of Section 4.4 as well. Therefore we can see that existence of solution of $(\overline{\underline{\mathcal{P}}}_{\underline{F}_z})$ cannot be showed by using the Direct Method of Calculus of Variations – optimality of $\underline{v}$ for this problem is by sheer construction.

Since we cannot rely on the aforementioned results directly we will recover them "by hand". We will concentrate on the system of PDEs of elasticity similar to the one in Corollary 4.24. We have already started this process in Lemma 6.14 where we showed that $\underline{v}$ as above is an element of $\mathrm{dom}(\underline{e}_{\underline{\mu}})$ for the domain $\mathscr{U} = \underline{\mathbb{R}}^d \setminus \underline{\Upsilon}_0$. We proceed to show that PDEs analogous to those in Theorem 4.22 hold true:

Theorem 6.23. *The quadruple $\underline{v}, \underline{\mu}, \underline{\sigma}, \underline{\mathscr{H}}$ together with the load $\underline{F}_z$ satisfies the following system of equations:*

$$\begin{cases} (i) & -\mathrm{Div}\big(\underline{\sigma}\underline{\mu}\big) = \underline{F}_z \quad \textit{in} \quad \underline{\mathbb{R}}^d \setminus \underline{\Upsilon}_0, \\ (ii) & \underline{\rho}_+\big(\underline{e}(\underline{v})\big) \le 1 \qquad \mathcal{L}^d\textit{-a.e. in } \underline{\Omega}, \\ (iii) & \underline{\mu}(\underline{\Upsilon}_0) = 0, \\ (iv) & \underline{\sigma} \in \partial j_{+,\underline{\mu}}\big(\underline{x},\, \underline{\mathscr{H}},\, \underline{e}_{\underline{\mu}}(\underline{v})\big) \qquad \underline{\mu}\textit{-a.e.}, \\ (v) & \mathrm{Tr}\,\underline{\sigma} = 1 \qquad \underline{\mu}\textit{-a.e.} \end{cases} \tag{6.39}$$

Proof. Let us start by recalling that, since $(u, w) \in \mathrm{Lip}_0(\Omega; \mathbb{R}^{d+1})$ and $(\vartheta, \varsigma) = (q\mu, \sigma\mu)$ are optimal for $(\overline{\mathscr{P}})$ and, respectively, $(\mathscr{P}^*)$, we know that the optimality conditions in (5.67) are met. The equation (i) follows by Lemma 6.17 since $\underline{\sigma}\underline{\mu} = \underline{\varsigma}$, while inequality (ii) was already established in (6.21). By definition of $\underline{\mu}$ we have $\underline{\mu}(\underline{\Upsilon}_0) = \sqrt{2}\,\mu(\partial\Omega)$ therefore condition (iii) in (6.39) follows from (iii) in (5.67). Condition (v) is (6.28) exactly.

To prove (iv) we write down a chain of equalities, staring with the $\underline{\mu}$-tangential integration by parts formula from Theorem 2.57 which holds thanks to (i) and the fact that $\underline{v} \in \mathrm{dom}(\underline{e}_{\underline{\mu}})$:

$$\int_{\underline{\mathbb{R}}^d \setminus \underline{\Upsilon}_0} \big\langle \underline{e}_{\underline{\mu}}(\underline{v}), \underline{\sigma} \big\rangle \, d\underline{\mu} = \int_{\underline{\mathbb{R}}^d} \langle \underline{v}, \underline{F}_z \rangle = \underline{Z}_{\underline{F}_z} = \int_{\underline{\mathbb{R}}^d} \mathrm{Tr}\,\underline{\varsigma} = \int_{\underline{\mathbb{R}}^d} \mathrm{Tr}\,\underline{\sigma}\, d\underline{\mu} = \int_{\underline{\mathbb{R}}^d} \underline{\rho}^0_{+,\underline{\mu}}(\underline{x}, \underline{\sigma})\, d\underline{\mu}$$

where the second and third equality is by optimality of $\underline{v}$ and $\underline{\varsigma}$, while the last equality acknowledges that $\underline{\sigma}(\underline{x}) \in \underline{\mathcal{S}}^{d\times d}_+ \cap M^e_{\underline{\mu}}(\underline{x})$ for $\underline{\mu}$-a.e. $\underline{x}$, cf. assertion (i) in Theorem 2.57. The last integral may be taken over $\underline{\mathbb{R}}^d \setminus \underline{\Upsilon}_0$ thanks to (iii). Then, since by Lemma 6.14 we have $\underline{\rho}_{+,\underline{\mu}}\big(\underline{x}, \underline{e}_{\underline{\mu}}(\underline{v})\big) \le 1$ $\underline{\mu}$-a.e., by comparing the first and the last integral above we infer that the extremality relation holds

$$\big\langle \underline{e}_{\underline{\mu}}(\underline{v}), \underline{\sigma} \big\rangle = \underline{\rho}^0_{+,\underline{\mu}}(\underline{x}, \underline{\sigma}) = \mathrm{Tr}\,\underline{\sigma} = 1 \qquad \underline{\mu}\text{-a.e.} \tag{6.40}$$

Let us now observe that for $\underline{\mu}$-a.e. $\underline{x}$ the prerequisites of Corollary 4.16 are met for the triple $\xi = \underline{e}_{\underline{\mu}}(\underline{v})(x)$, $\sigma = \underline{\sigma}(\underline{x})$, $\breve{H} = \underline{\mathscr{H}}(\underline{x})$. At the same time $\underline{\mathscr{H}}(\underline{x}) \in \underline{\mathcal{K}}_1\big(\underline{\sigma}(\underline{x})\big)$ for $\underline{\mu}$-a.e.

$\underline{x}$ by construction. Combining the latter facts with (6.40) furnishes the constitutive law (iv) in (6.39) through Corollary 4.16. □

Let us represent the optimal Hooke field $\underline{\lambda}$ through the elastic material distribution $\check{\underline{\mu}}$ and the anisotropy function $\check{\mathscr{H}}$ as below

$$\underline{\lambda} = \check{\mathscr{H}}\,\check{\underline{\mu}} \qquad \text{where} \qquad \check{\underline{\mu}} = \frac{C_0}{\sqrt{2}\,\mathcal{Z}}\,\underline{\mu}, \qquad \check{\mathscr{H}} = \mathscr{H}.$$

Then we propose the displacement and the relative stress functions

$$\check{\underline{v}}(\underline{x}) = \frac{\sqrt{2}\,\mathcal{Z}}{C_0}\Big(w(x) + \sqrt{2}\,u(x)\,\mathrm{e}_{\underline{d}}\Big), \qquad \check{\underline{\sigma}} = \frac{\sqrt{2}\,\mathcal{Z}}{C_0}\,\underline{\sigma}$$

We can see that $\check{\underline{\sigma}}\check{\underline{\mu}} = \underline{\sigma}\underline{\mu}$ therefore, by using a reasoning similar to the one above Corollary 4.24, from the system (6.39) we may deduce that:

Corollary 6.24. *The displacement function $\check{\underline{v}}$ and the (relative) stress function $\check{\underline{\sigma}}$ satisfy the partial differential equations of elasticity in the optimal form $\underline{\lambda} = \check{\mathscr{H}}\,\check{\underline{\mu}}$ subjected to the load $\underline{F}_z$ and pinned on the curve $\underline{\Upsilon}_0 = \partial\Omega \times \{0\}$, namely:*

(i) the equilibrium equation

$$-\mathrm{Div}\big(\check{\underline{\sigma}}\check{\underline{\mu}}\big) = \underline{F}_z \quad \textit{in} \quad \mathbb{R}^d \setminus \underline{\Upsilon}_0,$$

(ii) the constitutive law

$$\check{\underline{\sigma}} \in \partial j_{+,\check{\underline{\mu}}}\big(\underline{x}\,,\,\check{\mathscr{H}}\,,\,\underline{e}_{\check{\underline{\mu}}}(\check{\underline{v}})\big) \qquad \check{\underline{\mu}}\text{-a.e.} \tag{6.41}$$

Moreover, the strain and the relative stress are uniform in the following sense:

$$\underline{\rho}_{+,\check{\underline{\mu}}}\big(\underline{e}_{\check{\underline{\mu}}}(\check{\underline{v}})\big) = \tfrac{\sqrt{2}\,\mathcal{Z}}{C_0}, \qquad \mathrm{Tr}\,\check{\underline{\sigma}} = \tfrac{\sqrt{2}\,\mathcal{Z}}{C_0} \qquad \check{\underline{\mu}}\text{-a.e.}$$

By recalling that the elevation function was defined as $z = \frac{1}{\sqrt{2}}\,u$ we can extract an astonishingly simple property of the vertical displacement in the optimally designed form:

Corollary 6.25. *At every point of the form $\overline{S}_z$ the vertical component of the displacement is proportional to the elevation of this point, i.e.*

$$\big\langle \check{\underline{v}}\big(\underline{z}(x)\big), \mathrm{e}_{\underline{d}}\big\rangle = \tfrac{2\sqrt{2}\,\mathcal{Z}}{C_0}\,z(x) \qquad \forall\, x \in \overline{\Omega}.$$

6.3.3. Example of an optimal form

The developed method of constructing optimal forms will be now put to use on the example of a cylindrical symmetric problem, namely when $\underline{\Omega} = \Omega \times (-h_0, h_0) \subset \mathbb{R}^3$ is a cylinder of revolution and f is radial symmetric. Accordingly, we must turn our attention to the radial symmetric optimal membrane problem formulated on the disk Ω. The latter problem was fully solved in Section 5.4.2 where optimal functions (u, w) where proved to always be Lipschitz continuous, therefore we are in a position to derive universal formulas for the optimal cylindrical forms. Instead we will focus on the particular example when the load is uniformly distributed over Ω:

Example 6.26 (**An optimal cylindrical symmetric form**)**.** Let us consider the optimal form-finding problem for the data: $\Omega = \big\{|x| < R \,:\, x \in \mathbb{R}^2\big\}$, $f = p\,\mathcal{L}^2 \llcorner \Omega$ where $R, p > 0$ and $\underline{\Omega} = \Omega \times (-h_0, h_0)$ where $h_0 = R$. Following the construction in Section 6.3.1 we address Example 5.45 where the corresponding problems $(\overline{\mathscr{P}})$ and $(\mathscr{P}^*)$ were solved: the respective solutions are of the form $(u, w) = (\tilde{u}, w_r\,\mathrm{e}_r)$ and $(\vartheta, \varsigma) = (q_r\,\mathrm{e}_r\,,\, \sigma_{rr}\,\mathrm{e}_r \otimes \mathrm{e}_r)\,\mathcal{L}^2 \llcorner \Omega$ where

$$\tilde{u}(r) = \frac{\sqrt{10}}{3}\frac{R^3 - r^3}{R^2}, \quad w_r(r) = r - \frac{r^5}{R^4}, \quad q_r(r) = -\frac{p\,r}{2}, \quad \sigma_{rr}(r) = \frac{pR^2}{2\sqrt{10}}\,\frac{1}{r}. \tag{6.42}$$

In the sequel of this example let us agree that by writing $\underline{x}$, x, r in one equality (or a chain of equalities) we mean $x = \pi_{\mathbb{R}^2}(\underline{x})$ and $r = |x| = \langle x, \mathrm{e}_r\rangle$ or $r = |\pi_{\mathbb{R}^2}(\underline{x})| = \langle \underline{x}, \mathrm{e}_r\rangle$. In Fig. 6.2(a) the pre-stress ς is showed as a family of strings to which it may be decomposed.

We begin the construction in accordance with Section 6.3.1. For the elevation function we choose the smooth radial symmetric function $z \in C^\infty(\overline{\Omega}; \mathbb{R})$ as below

$$z(x) := \tilde{z}(r) := \frac{1}{\sqrt{2}}\,\tilde{u}(r) = \frac{\sqrt{5}}{3}\frac{R^3 - r^3}{R^2};$$

the graph $\overline{S}_z$ is displayed in Fig. 6.2(b). Next we define the measure $\mu \in \mathcal{M}_+(\overline{\Omega})$ that here happens to be absolutely continuous with respect to the Lebesgue measure:

$$\mu = \varrho^0(\vartheta, \varsigma) = m\,\mathcal{L}^2 \llcorner \Omega$$

where for a.e. x the function $m \in L^1(\Omega)$ reads

$$\begin{aligned} m(x) = \varrho^0\Big(q_r(r)\,\mathrm{e}_r\,,\, \sigma_{rr}(r)\,\mathrm{e}_r \otimes \mathrm{e}_r\Big) &= \sigma_{rr}(r) + \frac{1}{2}\,\frac{1}{\sigma_{rr}(r)}\,\big(q_r(r)\big)^2 = \frac{pR^2}{2\sqrt{10}\,r} + \frac{\sqrt{10}\,p\,r^3}{4R^2} \\ &= \sigma_{rr}(r)\left(1 + \tfrac{1}{2}\big(\tilde{u}'(r)\big)^2\right) = \sigma_{rr}(r)\left(1 + \big(\tilde{z}'(r)\big)^2\right). \end{aligned}$$

Next we compute the function $\sigma \in L^1(\Omega; \mathcal{S}^{2\times 2}_+)$:

$$\sigma(x) = \frac{d\varsigma}{d\mu}(x) = \frac{\sigma_{rr}(|x|)\,\mathrm{e}_r \otimes \mathrm{e}_r}{m(x)} = \frac{1}{1 + \big(\tilde{z}'(r)\big)^2}\;\mathrm{e}_r \otimes \mathrm{e}_r.$$

We are ready to define the 3D objects, starting with measures $\underline{\mu} = \sqrt{2}\,\underline{z}\,\#\,\mu$ and $\underline{F}_z = \mathrm{e}_3\,\underline{z}\,\#\,f$. To produce explicit formulas we may use the Area Formula (see [82, Section 3.3, Theorem 1]) thanks to which we have $\underline{z}\,\#\,\big(J_{\underline{z}}\,\mathcal{L}^2 \llcorner \Omega\big) = \mathcal{H}^2 \llcorner S_z$ where $J_{\underline{z}}$ is the Jacobian of the function $\underline{z} = (\mathrm{id}, z)$:

$$J_{\underline{z}}(x) := \left(\det\Big(\nabla\underline{z}(x)\,\big(\nabla\underline{z}(x)\big)^\top\Big)\right)^{\frac{1}{2}} = \sqrt{1 + |\nabla z(x)|^2} = \sqrt{1 + (\tilde{z}'(r))^2} = \sqrt{1 + 5\,\frac{r^4}{R^4}}.$$

If in addition we acknowledge the change of variable formula in [82, Section 3.3, Theorem 2] we obtain

$$\underline{\mu} = \left(\sqrt{2}\,\left(\frac{m}{J_{\underline{z}}}\right) \circ \underline{z}^{-1}\right)\mathcal{H}^2 \llcorner S_z \qquad \underline{F}_z = \left(p\,\mathrm{e}_3\left(\frac{1}{J_{\underline{z}}}\right) \circ \underline{z}^{-1}\right)\mathcal{H}^2 \llcorner S_z$$

and the density $\underline{m} \in L^1(S_z)$ of $\underline{\mu}$ with respect to $\mathcal{H}^2 \llcorner S_z$ may be computed as

$$\underline{m}(\underline{x}) = \sqrt{2}\,\sigma_{rr}(r)\,\sqrt{1 + (\tilde{z}'(r))^2} = \frac{pR^2}{2\sqrt{5}\,r}\,\sqrt{1 + 5\,\frac{r^4}{R^4}}. \tag{6.43}$$

To compute the function $\underline{\sigma} \in L^\infty(S_z; \mathcal{S}^{3\times 3}_+)$ we will use the formula (6.34) although, since $\underline{z}$ is differentiable, we may take $\nabla \underline{z}$ instead of $\nabla_\mu \underline{z}$. After simplifications, at a.e. $\underline{x}$

$$\underline{\sigma}(\underline{x}) = \Big[\mathrm{I} \quad \nabla z(x)\Big]^\top \sigma(x) \Big[\mathrm{I} \quad \nabla z(x)\Big] = \underline{\tau}_{S_z,r}(\underline{x}) \otimes \underline{\tau}_{S_z,r}(\underline{x}), \qquad \underline{\tau}_{S_z,r}(\underline{x}) := \frac{\mathrm{e}_r + \tilde{z}'(r)\,\mathrm{e}_3}{\sqrt{1+(\tilde{z}'(r))^2}}$$

where the unit vector $\underline{\tau}_{S_z,r}$ can be checked to lie in the intersection of the tangent space to S_z at $\underline{x}$ and the vertical plane containing $\underline{x}$ and the origin, thus supporting the observation in (6.35). Since $\underline{\sigma}$ is a.e. rank-one the criterion (6.30) turns trivial, i.e. necessarily

$$\underline{\mathscr{H}}(\underline{x}) = \underline{\tau}_{S_z,r}(\underline{x}) \otimes \underline{\tau}_{S_z,r}(\underline{x}) \otimes \underline{\tau}_{S_z,r}(\underline{x}) \otimes \underline{\tau}_{S_z,r}(\underline{x}).$$

The dome-like optimal form is displayed in Fig. 6.2(c): it is a graphical representation of the stress field $\underline{\varsigma} = \underline{\sigma}\,\underline{\mu}$ (or the Hooke field $\underline{\lambda} = \underline{\mathscr{H}}\,\underline{\mu}$) that decomposes to curved cables – note that in contrast to ς the cables are of varying cross-sectional areas, which is due to the square root factor in (6.43).

Instead of computing the elastic deformation $\check{\underline{v}}$ we will study the rescaled function $\underline{v} = \frac{C_0}{\sqrt{2}\,\mathcal{Z}}\,\check{\underline{v}}$ which solves $(\overline{\mathcal{P}}_{\underline{F}_z})$:

$$\underline{v}(\underline{x}) = w_r(r)\,\mathrm{e}_r + \sqrt{2}\,\tilde{u}(r)\,\mathrm{e}_3.$$

By using the standard formulas for calculus in the polar coordinate system we obtain the strain tensor function

$$\underline{e}(\underline{v})(\underline{x}) = w_r'(r)\,\mathrm{e}_r \otimes \mathrm{e}_r + \frac{w_r(r)}{r}\,\mathrm{e}_\theta \otimes \mathrm{e}_\theta + \sqrt{2}\,\tilde{u}'(r)\,\mathrm{e}_r \overset{s}{\otimes} \mathrm{e}_3.$$

We can compute the component of the strain along the material fibres, i.e.

$$\begin{aligned}\Big\langle \underline{e}(\underline{v})(\underline{x})\,,\, \underline{\tau}_{S_z,r}(\underline{x}) \otimes \underline{\tau}_{S_z,r}(\underline{x})\Big\rangle &= \frac{1}{1+(\tilde{z}'(r))^2}\Big(w_r'(r) + \sqrt{2}\,\tilde{u}'(r)\,\tilde{z}'(r)\Big)\\ &= \frac{1}{1+(\tilde{z}'(r))^2}\bigg(\Big(\tfrac{1}{2}\big(\tilde{u}'(r)\big)^2 + w_r'(r)\Big) + \big(\tilde{z}'(r)\big)^2\bigg) = 1\end{aligned}$$

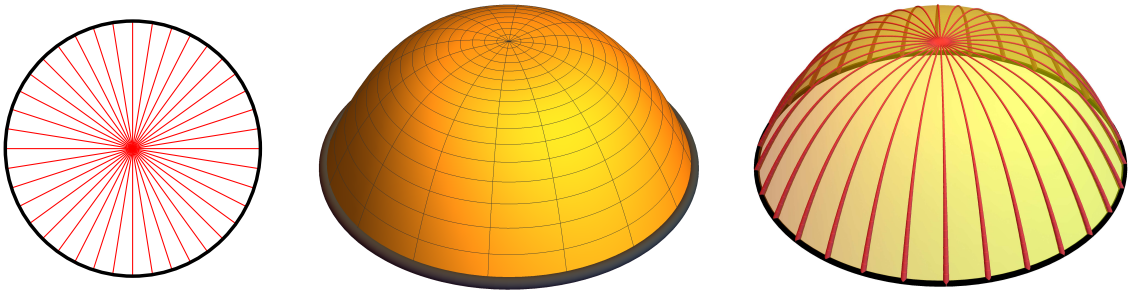
where after manipulating with the relation $\tilde{z}(r) = \frac{1}{\sqrt{2}}\,\tilde{u}(r)$ we have exploited the optimality condition (iv') in (5.73). As a result we can check that the linear Hooke law is met:

$$\underline{\mathscr{H}}(\underline{x})\Big(\underline{e}(\underline{v})(\underline{x})\Big) = \underline{\tau}_{S_z,r}(\underline{x}) \otimes \underline{\tau}_{S_z,r}(\underline{x}) = \underline{\sigma}(\underline{x})$$

therefore $\underline{\sigma} \in \partial j\big(\underline{\mathscr{H}}, \underline{e}(\underline{v})\big)$ since j is quadratic, see (6.1). The constitutive law $\underline{\sigma} \in \partial j_+\big(\underline{\mathscr{H}}, \underline{e}(\underline{v})\big)$ must hold as well by virtue of Corollary 3.21. △

Remark 6.27. In the example above we have used the Area Formula to produce explicit formula for the 3D measure $\underline{\mu} = \sqrt{2}\,\underline{z}\,\#\,\mu$. This was possible only because $\mu \ll \mathcal{L}^2$ in this particular example. Very often (assuming arbitrary d and $\underline{d} = d+1$) the measure μ will be of the form $\mu = \sum_{k=1}^d m_k\,\mathcal{H}^k \llcorner E_k$ where, for each k, E_k is a k-rectifiable set and $m_k \in L^1_{\mathcal{H}^k}(E_k)$. Then we may employ the *Generalized Area Formula* that may be found in [8, Section 2.11, Theorem 2.91] to arrive at (recall that $\underline{z}$ is Lipschitz continuous and injective)

$$\underline{\mu} = \sqrt{2}\,\underline{z}\,\#\,\mu = \sum_{k=1}^d \left(\sqrt{2}\left(\frac{m_k}{J^{E_k}_{\underline{z}}}\right)\circ \underline{z}^{-1}\right)\mathcal{H}^k \llcorner \big(\underline{z}(E_k)\big)$$



(a) (b) (c)

Figure 6.2: The cylindrical symmetric optimal form-finding problem for $f = p\,\mathcal{L}^2 \llcorner \Omega$: (a) the 2D pre-stress ς solving the optimal membrane problem $(\mathscr{P}^*)$; (b) the graph $\overline{S}_z$ on which the optimal form concentrates; (c) the optimal dome-like form represented by the stress field $\underline{\varsigma}$ obtained by "unprojection" of ς onto $\overline{S}_z$.

where for $\mathcal{H}^k$-a.e. $x \in E_k$ the number $J^{E_k}_{\underline{z}}(x)$ is the k-dimensional Jacobian computed for the suitably defined derivative of $\underline{z}$ that is tangential with respect to E_k. For the details the reader is referred to [8]. Here it suffices to say that if for each k the set E_k is a finite union of k-dimensional C^2 manifolds and $C_1 \leq m_k \leq C_2$ for some $C_1, C_2 > 0$ then $J^{E_k}_{\underline{z}}(x) = \sqrt{1 + |\nabla_\mu z(x)|^2}$ for μ-a.e. x thanks to Proposition 2.59 and Remark 2.61. For the time being it seems that for general Radon measures μ one is bound to work with the abstract push-forward measure.

6.4. Optimal grid-shells constructed from optimal string systems

The construction of optimal forms from the previous section relied on the regularity assumption on the solution of optimal membrane problem: the functions (u, w) had to be Lipschitz continuous. Now we will be able to drop this assumption, but in exchange for other one: we shall assume that there exists a string system $(\pi, \mathscr{S})$ that solves the problem $(\widetilde{\mathscr{P}}^*)$, see Section 5.5. Although it is very easy to come up with examples when this existence fails – see Remark 5.59 – it appears as if optimal string systems exist whenever the load f is finitely supported, cf. Proposition 5.107.

The construction of the optimal form itself will be easier than in Section 6.3: the d-dimensional string system $(\pi, \mathscr{S}) \in \mathcal{M}(\overline{\Omega} \times \overline{\Omega}; \mathbb{R} \times \mathbb{R}_+)$ will be "elevated" to a $\underline{d}$-dimensional truss in tension being a pairing measure $\underline{\mathscr{S}} \in \mathcal{M}(\overline{S}_z \times \overline{S}_z; \mathbb{R}_+)$ where again $z = \frac{1}{\sqrt{2}}\,u$. Then, together with $\underline{F} = \underline{F}_z$, the stress field $\underline{\varsigma} = \underline{\varsigma}_{\underline{\mathscr{S}}}$ (we use an integration formula as in (4.58)) will be proved to solve the Prager problem (PPP). This is, however, only half of success since the fact that the joints of the truss bars lie on $\overline{S}_z$, namely $\operatorname{spt} \mathscr{S} \subset \overline{S}_z \times \overline{S}_z$, does not guarantee that the bars lie on this surface as well. In other words we must prove that $\operatorname{spt} \underline{\varsigma}_{\underline{\mathscr{S}}} \subset \overline{S}_z$ in order to show that the obtained structure is a form and not a truly spacial structure. Such class of forms – that are composed of truss bars – will be referred to as *grid-shells*.

There is another motivation for constructing optimal grid-shells out of optimal string systems: it leads to the numerical method for the optimal form finding problems (PFF), (EFF) and for the Prager problems (PPP), (EPP) all at once, see Section 6.5 for more details.

6.4.1. The construction of the grid-shell

In the whole Section 6.4 the bounded domain $\Omega \subset \mathbb{R}^d$ will be assumed to be convex, although this condition is merely for convenience and could be dropped, see Remark 5.52. We choose any load $f \in \mathcal{M}(\overline{\Omega};\mathbb{R})$ and a height parameter $h_0 \geq \frac{1}{2}\operatorname{diam}(\Omega)$ furnishing a convex $\underline{d}$-dimensional domain $\underline{\Omega} = \Omega \times (-h_0, h_0)$. Straightaway (prior to the construction unlike in Section 6.3) we make

$$\text{an assumption:} \qquad \text{the problem } (\widetilde{\mathscr{P}^*}) \text{ admits a solution } (\pi, \mathscr{S}). \tag{6.44}$$

Throughout the rest of Section 6.4 we choose and fix two pairs

$$(u,w) \in \overline{\mathcal{X}}_1 \quad \text{and} \quad (\pi,\mathscr{S}) \in \mathcal{M}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}\times\mathbb{R}_+) \qquad \text{that solve} \qquad (\overline{\mathscr{P}}) \ \text{ and } \ (\widetilde{\mathscr{P}^*}),$$

respectively. Without the loss of generality we may assume that $\mathscr{S}(\Delta) = 0$ where Δ is the diagonal of $\overline{\Omega}\times\overline{\Omega}$. We must strongly emphasize that the assumption (6.31) on the Lipschitz continuity of (u,w) is now dropped and the pair is any element of $\overline{\mathcal{X}}_1$. In particular $u \in C^{0,\frac{1}{2}}(\Omega) \cap W^{1,2}(\Omega)$ while $w \in BV \cap L^\infty(\Omega;\mathbb{R}^d)$. At the same time $w \in \mathcal{W}_\Omega$ (see Section 5.3.2) and thus $\mathbf{v} = \mathrm{id} - \mathbf{i}^{-1}(w) \in \mathbf{V}_\Omega$ is a maximal monotone map, while $(u,\mathbf{v})$ solves the geometric problem $(\mathscr{P}_{\mathrm{geo}})$ by Proposition 5.84. In turn, by virtue of Theorem 5.95 the two pairs $(u,\mathbf{v})$ and $(\pi,\mathscr{S})$ solve the system of optimality conditions (5.147), including the one that would be the most important for the truss constructed below:

$$\begin{cases} \pi = \alpha\mathscr{S} \quad \text{for} \quad \alpha \in L^1_{\mathscr{S}}(\overline{\Omega}\times\overline{\Omega};\mathbb{R}), \\ \alpha(x,y) = \frac{u(y)-u(x)}{|y-x|} \qquad \text{for } \mathscr{S}\text{-a.e. } (x,y) \in \overline{\Omega}\times\overline{\Omega}. \end{cases} \tag{6.45}$$

The construction of the grid-shell starts with the elevation function $z \in C_0(\Omega)$ that is defined identically as in Section 6.3:

$$z := \frac{1}{\sqrt{2}}\,u, \qquad \underline{z}(x) := \big(x, z(x)\big) = x + z(x)\,\mathrm{e}_{\underline{d}}.$$

By pushing-forward we construct a pairing measure that represents a $\underline{d}$-dimensional generalized truss (in the sense of Section 4.7.1) in pure tension:

$$\underline{\mathscr{S}} := (\underline{z},\underline{z}) \,\#\, \Big(\sqrt{2}\,\big(\chi^*_{\mathrm{K}}(\alpha,1)\big)^{\frac{1}{2}}\mathscr{S}\Big) \tag{6.46}$$

where $\chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big) = 1 + \frac{1}{2}\big(\alpha(x,y)\big)^2$, see the formula (5.83). The function $\chi^*_{\mathrm{K}}(\alpha,1)$ is $\mathscr{S}$-integrable therefore $\underline{\mathscr{S}}$ is a push-forward of a Radon measure on $\overline{\Omega}\times\overline{\Omega}$ by a continuous function $(\underline{z},\underline{z})$. The measure $\underline{\mathscr{S}}$ is thus a Radon measure on $\overline{\underline{\Omega}}\times\overline{\underline{\Omega}}$; it is clearly supported on $\overline{S}_z \times \overline{S}_z$ hence we may write $\underline{\mathscr{S}} \in \mathcal{M}_+(\overline{S}_z\times\overline{S}_z)$.

Remark 6.28. Let us observe that the mapping $(\underline{z},\underline{z}) : \overline{\Omega}\times\overline{\Omega} \to \overline{S}_z\times\overline{S}_z$ is a homeomorphism. Considering the definition of the push-forward and the fact that both $\chi^*_{\mathrm{K}}(\alpha,1)$ and $1/\big(\chi^*_{\mathrm{K}}(\alpha,1)\big)$ are $\mathscr{S}$-integrable we can deduce that a set $N \subset \overline{\Omega}\times\overline{\Omega}$ is $\mathscr{S}$-negligible if

and only if the set $(\underline{z},\underline{z})(N)$ is $\mathscr{S}$-negligible. Consequently, a property $P(\underline{x},\underline{y})$ holds true for $\underline{\mathscr{S}}$-a.e. $(\underline{x},\underline{y})$ if and only if the property $P\big(\underline{z}(x),\underline{z}(y)\big)$ is met for $\mathscr{S}$-a.e. (x,y).

The distance between points lying on the graph $\overline{S}_z$ may be computed using Pythagorean theorem and further rewritten (assuming that $x \neq y$):

$$|\underline{z}(y)-\underline{z}(x)|^2 = |y-x|^2 + |z(y)-z(x)|^2 = |y-x|^2\left(1+\frac{1}{2}\Big(\frac{u(y)-u(x)}{|y-x|}\Big)^2\right).$$

We can now exploit the optimality condition (6.45) to observe that

$$|\underline{z}(y)-\underline{z}(x)| = |y-x|\left(\chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big)\right)^{\frac{1}{2}} \qquad \text{for } \mathscr{S}\text{-a.e. } (x,y) \tag{6.47}$$

which essentially furnishes a formula for the lengths of bars of the truss $\underline{\mathscr{S}}$.

Remark 6.29. Based on (6.47) we see that the value $\big(\chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big)\big)^{\frac{1}{2}}$ is the slope of a bar $[\underline{z}(x),\underline{z}(y)] \subset \mathbb{R}^{\underline{d}}$, i.e. the tangent of the angle between the bar and the base plane $\mathbb{R}^d$. This sheds a new light on the formula (6.46) for the $\underline{d}$-dimensional truss $\underline{\mathscr{S}}$. The horizontal d-dimensional strings $[x,y]$ of cross-sectional are $\mathscr{S}(dxdy)$ are elevated to $\underline{d}$-dimensional bars $[\underline{x},\underline{y}] = [\underline{z}(x),\underline{z}(y)]$ while their cross-sectional area $\underline{\mathscr{S}}(d\underline{x}d\underline{y})$ is increased by the factor of the slope $|\underline{z}(y)-\underline{z}(x)|/|y-x|$ (up to neglecting the factor $\sqrt{2}$).

If to the truss $\underline{\mathscr{S}}$ we apply the load transmitted by elevation, i.e.

$$\underline{F}_z := \mathrm{e}_{\underline{d}}\,\underline{z}\,\#\,f$$

then equilibrium of the truss (in the sense of Section 4.7.1) can be showed along with computing the volume of the truss:

Lemma 6.30. *For any $\underline{\phi} \in C_0\big(\mathbb{R}^d \setminus \underline{\Upsilon}_0;\mathbb{R}^{\underline{d}}\big)$ there holds an equality*

$$\int_{\overline{\Omega}\times\overline{\Omega}} \big\langle \underline{\phi}(\underline{y})-\underline{\phi}(\underline{x}),\underline{\tau}^{\underline{x},\underline{y}}\big\rangle\, \underline{\mathscr{S}}(d\underline{x}d\underline{y}) = \int_{\overline{\Omega}} \langle \underline{\phi},\underline{F}_z\rangle. \tag{6.48}$$

where $\underline{\tau}^{\underline{x},\underline{y}} := \frac{\underline{y}-\underline{x}}{|\underline{y}-\underline{x}|}$. Moreover we have

$$\int_{\overline{\Omega}\times\overline{\Omega}} \big|\underline{y}-\underline{x}\big|\, \underline{\mathscr{S}}(d\underline{x}d\underline{y}) = \sqrt{2}\,\mathcal{Z}.$$

Proof. By using the change of variable formula for the push-forward we obtain

$$\begin{aligned}\int_{\overline{\Omega}\times\overline{\Omega}} \big|\underline{y}-\underline{x}\big|\, \underline{\mathscr{S}}(d\underline{x}d\underline{y}) &= \int_{\overline{\Omega}\times\overline{\Omega}} |\underline{z}(y)-\underline{z}(x)|\,\sqrt{2}\left(\chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big)\right)^{\frac{1}{2}}\mathscr{S}(dxdy)\\ &= \sqrt{2}\int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\,\chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big)\,\mathscr{S}(dxdy)\\ &= \sqrt{2}\int_{\overline{\Omega}\times\overline{\Omega}} |y-x|\,\chi^*_{\mathrm{K}}(\pi,\mathscr{S})(dxdy) = \sqrt{2}\,\inf\widetilde{\mathscr{P}^*} = \sqrt{2}\,\mathcal{Z}\end{aligned}$$

where to pass to the second line we used (6.47). Similarly we compute

$$\text{LHS of (6.48)} = \int_{\overline{\Omega}\times\overline{\Omega}} \Big\langle \underline{\phi}\big(\underline{z}(y)\big)-\underline{\phi}\big(\underline{z}(x)\big),\frac{\underline{z}(y)-\underline{z}(x)}{|\underline{z}(y)-\underline{z}(x)|}\Big\rangle\sqrt{2}\left(\chi^*_{\mathrm{K}}\big(\alpha(x,y),1\big)\right)^{\frac{1}{2}}\mathscr{S}(dxdy).$$

Let us denote the functions $\phi \in C_0(\Omega;\mathbb{R}^d)$, $\varphi \in C_0(\Omega)$ such that $\phi(x) + \varphi(x)\,\mathrm{e}_{\underline{d}} = \underline{\phi}\big(\underline{z}(x)\big)$. Then we have

$$\Big\langle \underline{\phi}\big(\underline{z}(y)\big) - \underline{\phi}\big(\underline{z}(x)\big), \underline{z}(y) - \underline{z}(x)\Big\rangle = \Big\langle \phi(y) - \phi(x), y - x\Big\rangle + \frac{1}{\sqrt{2}}\,\big(\varphi(y) - \varphi(x)\big)\big(u(y) - u(x)\big)$$

and, after using (6.47) again and by considering that $\frac{u(y)-u(x)}{|y-x|} = \alpha(x,y)$ $\mathscr{S}$-a.e. owing to the optimality condition (iv) in (5.147), the left hand side of (6.48) can be further rewritten as

$$\begin{aligned}&\sqrt{2}\int_{\overline{\Omega}\times\overline{\Omega}} \Big\langle \phi(y) - \phi(x), \frac{y-x}{|y-x|}\Big\rangle \mathscr{S}(dxdy) + \int_{\overline{\Omega}\times\overline{\Omega}} \big(\varphi(y) - \varphi(x)\big)\,\alpha(x,y)\,\mathscr{S}(dxdy)\\ &= 0 + \int_{\overline{\Omega}\times\overline{\Omega}} \big(\varphi(y) - \varphi(x)\big)\,\pi(dxdy) = \int_{\overline{\Omega}} \varphi(x)\,f(dx) = \int_{\underline{\overline{\Omega}}} \big\langle \underline{\phi}(\underline{x}), \mathrm{e}_{\underline{d}}\big\rangle\,(\underline{z}\#f)(d\underline{x})\end{aligned}$$

where along the way we have exploited the fact that $(\pi,\mathscr{S}) \in \Xi(\Omega,f)$. The last integral may be readily recognized as $\int_{\underline{\overline{\Omega}}} \langle \underline{\phi}, \underline{F}_z\rangle$ by the very definition of $\underline{F}_z$. □

We are now in a position to put forward a pair $(\underline{\varsigma},\underline{F})$ being a candidate for solving the Prager problem (PPP) as well as the form-finding problem (PFF):

$$\underline{\varsigma} := \underline{\varsigma}_{\underline{\mathscr{S}}}, \qquad \underline{F} := \underline{F}_z$$

where the measure $\underline{\varsigma}_{\underline{\mathscr{S}}} \in \mathcal{M}(\overline{\Omega};\mathcal{S}^{\underline{d}\times\underline{d}}_+)$ is defined via the integration formula

$$\big\langle \underline{\psi}, \underline{\varsigma}_{\underline{\mathscr{S}}}\big\rangle_{(C,\mathcal{M})} = \int_{\overline{S}_z\times\overline{S}_z} \left(\int_{\underline{\overline{\Omega}}} \big\langle \underline{\psi}, \underline{\varsigma}^{\underline{x},\underline{y}}\big\rangle\right) \underline{\mathscr{S}}(d\underline{x}d\underline{y}) \qquad \forall\,\underline{\psi} \in C(\overline{\Omega};\mathcal{S}^{\underline{d}\times\underline{d}}) \tag{6.49}$$

where for $(\underline{x},\underline{y}) \in \overline{\Omega}\times\overline{\Omega}\setminus\Delta$ as in the previous chapters we put $\underline{\varsigma}^{\underline{x},\underline{y}} := \underline{\tau}^{\underline{x},\underline{y}} \otimes \underline{\tau}^{\underline{x},\underline{y}}\,\mathcal{H}^1\llcorner[\underline{x},\underline{y}]$.

As will appear below, with Lemma 6.30 at our disposal proving optimality of $(\underline{\varsigma}_{\underline{\mathscr{S}}},\underline{F}_z)$ for (PPP) will be straightforward. The same cannot be told about the form-finding problem (PFF), in other words: it is not trivial that $\operatorname{spt}\underline{\varsigma}_{\underline{\mathscr{S}}}$ is a subset of $\overline{S}_z$, although we know that $\operatorname{spt}\underline{\mathscr{S}} \subset \overline{S}_z\times\overline{S}_z$. It is simply due to the non-convexity of the graph $\overline{S}_z$ therefore it is not true that segments $[\underline{z}(x),\underline{z}(y)]$ are contained in $\overline{S}_z$ for arbitrary $(x,y) \in \overline{\Omega}\times\overline{\Omega}$. However, it is enough that such property is met for $\mathscr{S}$-a.e. (x,y), which we will be able to prove by employing results that were prepared in Chapter 5.

Theorem 6.31. *The pair $(\underline{\varsigma}_{\underline{\mathscr{S}}},\underline{F}_z)$ solves the plastic setting of the Prager problem* (PPP). *Moreover*

$$\operatorname{spt}\underline{\varsigma}_{\underline{\mathscr{S}}} \;\subset\; \overline{S}_z$$

and, as a result, the pair $(z,\underline{\varsigma}_{\underline{\mathscr{S}}})$ solves the plastic variant of the optimal form-finding problem (PFF).

Proof. The equilibrium equation $-\operatorname{Div}\underline{\varsigma}_{\underline{\mathscr{S}}} = \underline{F}_z$ in $\mathbb{R}^d\setminus\underline{\Upsilon}_0$ is due to condition (6.48) in Lemma 6.30, see Proposition 4.54. With $\underline{\mathrm{I}}$ being the $\underline{d}\times\underline{d}$ identity matrix we compute

$$\int_{\underline{\overline{\Omega}}} \operatorname{Tr}\underline{\varsigma}_{\underline{\mathscr{S}}} = \int_{\underline{\overline{\Omega}}} \big\langle \underline{\mathrm{I}}, \underline{\varsigma}_{\underline{\mathscr{S}}}\big\rangle = \int_{\overline{\Omega}\times\overline{\Omega}} \left(\int_{[\underline{x},\underline{y}]} \big\langle \underline{\mathrm{I}}, \underline{\varsigma}^{\underline{x},\underline{y}}\big\rangle\right) \underline{\mathscr{S}}(d\underline{x}d\underline{y}) = \int_{\overline{\Omega}\times\overline{\Omega}} \big|\underline{y} - \underline{x}\big|\,\underline{\mathscr{S}}(d\underline{x}d\underline{y}). \tag{6.50}$$

The latter integral is equal to $\sqrt{2}\,\mathcal{Z}$ by the "moreover part" of Lemma 6.30 hence the optimality of the pair $(\underline{\varsigma}_{\underline{\mathscr{S}}},\underline{F}_z)$ thanks to Theorem 6.9.

To show that $\operatorname{spt}\underline{\varsigma}_{\underline{\mathscr{S}}} \subset \overline{S}_z$ we use the statement (2) in the "moreover part" of Theorem 5.95. From it we deduce that for $\mathscr{S}$-a.e. pair (x,y) we have: $\big[\big(x,u(x)\big),\big(y,u(y)\big)\big] = \big\{\big(\tilde{x},u(\tilde{x})\big) : \tilde{x}\in[x,y]\big\}$. Considering that $\underline{z}(x) = \big(x, \frac{1}{\sqrt{2}}\,u(x)\big)$ we deduce that

$$[\underline{z}(x),\underline{z}(y)] \quad\subset\quad \overline{S}_z \qquad\qquad \text{for } \mathscr{S}\text{-a.e. } (x,y) \tag{6.51}$$

and therefore $[\underline{x},\underline{y}]\subset\overline{S}_z$ for any for $\underline{\mathscr{S}}$-a.e. $(\underline{x},\underline{y})$, cf. Remark 6.28. Consequently, if we take a test function $\underline{\psi}\in C(\overline{\Omega};\mathcal{S}^{\underline{d}\times\underline{d}})$ such that $\operatorname{spt}\underline{\psi}\subset\mathbb{R}^{\underline{d}}\setminus\overline{S}_z$, then $\int\langle\psi,\underline{\varsigma}^{\underline{x},\underline{y}}\rangle = 0$ for $\underline{\mathscr{S}}$-a.e. $(\underline{x},\underline{y})$ and thus $\big\langle\underline{\psi},\underline{\varsigma}_{\underline{\mathscr{S}}}\big\rangle_{(C,\mathcal{M})} = 0$. □

The optimal stress field $\underline{\varsigma}_{\underline{\mathscr{S}}}$ decomposes to bars making up the truss $\underline{\mathscr{S}}$ and at the same time it concentrates on a single surface $\overline{S}_z$. In other words we end up with a form that is a $\underline{d}$-dimensional truss. Such truss $\underline{\mathscr{S}}$ or the form $\underline{\varsigma}_{\underline{\mathscr{S}}}$ itself deserves to be called a (plastic) *optimal grid-shell*.

Similarly as in Section 6.3 the Hooke field $\underline{\lambda}$ that together with $\underline{F}=\underline{F}_z$ solves the elastic problems (EPP) and (EFF) could be easily constructed based on the field $\underline{\varsigma}=\underline{\varsigma}_{\underline{\mathscr{S}}}$ via Proposition 6.5. This way, however, one loses the information on the grid-shell structure of $\underline{\lambda}$. Instead we can define $\underline{\lambda}=\underline{\lambda}_{\underline{\mathscr{S}}}$ through another integration formula:

$$\big\langle\underline{\Psi},\underline{\lambda}_{\underline{\mathscr{S}}}\big\rangle_{(C,\mathcal{M})} = \frac{C_0}{\sqrt{2}\mathcal{Z}}\int_{\overline{S}_z\times\overline{S}_z}\left(\int_{\underline{\Omega}}\langle\underline{\Psi},\underline{\lambda}^{\underline{x},\underline{y}}\rangle\right)\underline{\mathscr{S}}(d\underline{x}d\underline{y}) \qquad \forall\,\underline{\Psi}\in C\big(\overline{\Omega};\mathscr{L}_+(\mathcal{S}^{\underline{d}\times\underline{d}})\big). \tag{6.52}$$

where $\underline{\lambda}^{\underline{x},\underline{y}} := \underline{\tau}^{\underline{x},\underline{y}}\otimes\underline{\tau}^{\underline{x},\underline{y}}\otimes\underline{\tau}^{\underline{x},\underline{y}}\otimes\underline{\tau}^{\underline{x},\underline{y}}\;\mathcal{H}^1\llcorner[\underline{x},\underline{y}]$. Since $\underline{\lambda}^{\underline{x},\underline{y}}\in\mathcal{M}(\overline{\Omega};\mathcal{K}_{fib})$ for each $(\underline{x},\underline{y})\in\overline{\Omega}\times\overline{\Omega}$ then also $\underline{\lambda}_{\underline{\mathscr{S}}}\in\mathcal{M}(\overline{\Omega};\mathcal{K}_{fib})$ by assertion (ii) of Theorem 3.35. On top of that, through a computation similar to (6.50) we may easily compute that $\int_{\overline{\Omega}}\operatorname{Tr}\underline{\lambda}_{\underline{\mathscr{S}}} = \int_{\overline{\Omega}}\big\langle\mathrm{Id},\underline{\lambda}_{\underline{\mathscr{S}}}\big\rangle = C_0$ where $\mathrm{Id}\in\mathscr{L}(\mathcal{S}^{\underline{d}\times\underline{d}})$ is the identity operator. All in all, the pair $(\underline{\lambda}_{\underline{\mathscr{S}}},\underline{F}_z)$ is admissible for (EPP) and we may move on to prove that $\underline{\lambda}_{\underline{\mathscr{S}}}$ is an optimal elastic grid-shell:

Theorem 6.32. *The pair $(\underline{\lambda}_{\underline{\mathscr{S}}},\underline{F}_z)$ solves the elastic setting of the Prager problem* (EPP). *Moreover*

$$\operatorname{spt}\underline{\lambda}_{\underline{\mathscr{S}}} \;\subset\; \overline{S}_z$$

and, as a result, the pair $(z,\underline{\lambda}_{\underline{\mathscr{S}}})$ solves the elastic variant of the optimal form-finding problem (EFF).

Proof. Let us consider any smooth displacement function $\underline{v}\in\mathcal{D}\big(\mathbb{R}^{\underline{d}}\setminus\underline{\Upsilon}_0;\mathbb{R}^{\underline{d}}\big)$. By employing Theorem 3.35 and then by using equality (5.36) we deliver a lower bound:

$$\begin{aligned}\int_{\underline{\Omega}} j_+\big(\underline{\lambda}_{\underline{\mathscr{S}}},\underline{e}(\underline{v})\big) &\geq \int_{\overline{\Omega}\times\overline{\Omega}}\frac{C_0}{\sqrt{2}\mathcal{Z}}\left(\int_{\underline{\Omega}} j_+\big(\underline{\lambda}^{\underline{x},\underline{y}},\underline{e}(\underline{v})\big)\right)\underline{\mathscr{S}}(d\underline{x}d\underline{y})\\ &= \int_{\overline{\Omega}\times\overline{\Omega}}\frac{C_0}{2\sqrt{2}\mathcal{Z}}\left(\int_{[\underline{x},\underline{y}]}\Big(\big(\langle\underline{e}(\underline{v}),\underline{\tau}^{\underline{x},\underline{y}}\otimes\underline{\tau}^{\underline{x},\underline{y}}\rangle\big)_+\Big)^2\,d\mathcal{H}^1\right)\underline{\mathscr{S}}(d\underline{x}d\underline{y}).\end{aligned}$$

where a_+ for $a \in \mathbb{R}$ stands for the positive part. Since $\underline{\varsigma}_{\underline{\mathscr{S}}}$ satisfies the equilibrium equation $-\mathrm{Div}\,\underline{\varsigma}_{\underline{\mathscr{S}}} = \underline{F}_z$ in $\mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0$ we obtain

$$\int_{\underline{\Omega}} \langle \underline{v}, \underline{F}_z \rangle = \int_{\underline{\Omega}} \big\langle \underline{e}(\underline{v}), \underline{\varsigma}_{\underline{\mathscr{S}}} \big\rangle = \int_{\underline{\Omega}\times\underline{\Omega}} \left(\int_{[\underline{x},\underline{y}]} \langle \underline{e}(\underline{v}), \underline{\tau}^{\underline{x},\underline{y}} \otimes \underline{\tau}^{\underline{x},\underline{y}} \rangle \, d\mathcal{H}^1 \right) \underline{\mathscr{S}}(d\underline{x}d\underline{y})$$
$$\leq \int_{\underline{\Omega}\times\underline{\Omega}} \left(\int_{[\underline{x},\underline{y}]} \big(\langle \underline{e}(\underline{v}), \underline{\tau}^{\underline{x},\underline{y}} \otimes \underline{\tau}^{\underline{x},\underline{y}} \rangle \big)_+ \, d\mathcal{H}^1 \right) \underline{\mathscr{S}}(d\underline{x}d\underline{y}).$$

Eventually we have

$$\int_{\underline{\Omega}} \langle \underline{v}, \underline{F}_z \rangle - \int_{\underline{\Omega}} j_+\big(\underline{\lambda}_{\underline{\mathscr{S}}}, \underline{e}(\underline{v})\big) \tag{6.53}$$
$$\leq \int_{\underline{\Omega}\times\underline{\Omega}} \left(\int_{[\underline{x},\underline{y}]} \left(\big(\langle \underline{e}(\underline{v}), \underline{\tau}^{\underline{x},\underline{y}} \otimes \underline{\tau}^{\underline{x},\underline{y}} \rangle \big)_+ - \frac{C_0}{2\sqrt{2}\mathcal{Z}} \left(\big(\langle \underline{e}(\underline{v}), \underline{\tau}^{\underline{x},\underline{y}} \otimes \underline{\tau}^{\underline{x},\underline{y}} \rangle \big)_+ \right)^2 \right) d\mathcal{H}^1 \right) \underline{\mathscr{S}}(d\underline{x}d\underline{y})$$
$$\leq \int_{\underline{\Omega}\times\underline{\Omega}} \left(\int_{[\underline{x},\underline{y}]} \frac{\mathcal{Z}}{\sqrt{2}C_0} \, d\mathcal{H}^1 \right) \underline{\mathscr{S}}(d\underline{x}d\underline{y}) = \frac{\mathcal{Z}}{\sqrt{2}C_0} \int_{\underline{\Omega}\times\underline{\Omega}} \big|\underline{y} - \underline{x}\big| \, \underline{\mathscr{S}}(d\underline{x}d\underline{y}) = \frac{\mathcal{Z}^2}{C_0}$$

where to pass to the last line we acknowledged that $\max_{t\in\mathbb{R}} \big\{ t - \frac{C_0}{2\sqrt{2}\mathcal{Z}} t^2 \big\} = \frac{\mathcal{Z}}{\sqrt{2}C_0}$ and the maximum is found for $\bar{t} = \frac{\sqrt{2}\mathcal{Z}}{C_0}$. By taking the supremum of the left hand side with respect to all $\underline{v} \in \mathcal{D}\big(\mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0; \mathbb{R}^{\underline{d}}\big)$ we deduce that $\underline{\mathcal{C}}(\underline{\lambda}_{\underline{\mathscr{S}}}, \underline{F}_z) \leq \frac{\mathcal{Z}^2}{C_0}$. Optimality of the pair $(\underline{\lambda}_{\underline{\mathscr{S}}}, \underline{F}_z)$ for (EPP) follows by Theorem 6.9. To prove that $\mathrm{spt}\,\underline{\lambda}_{\underline{\mathscr{S}}} \subset \overline{S}_z$ we use the same arguments as in the proof of Theorem 6.18. □

6.4.2. Analysis of the optimal elastic grid-shell – computing the deformation

For the optimal elastic grid-shell $\underline{\lambda}_{\underline{\mathscr{S}}}$ constructed in the previous subsection let us put:

$$\check{\underline{\mu}} := \mathrm{Tr}\,\underline{\lambda}_{\underline{\mathscr{S}}}, \qquad \check{\underline{\mathscr{H}}} := \frac{d\underline{\lambda}_{\underline{\mathscr{S}}}}{d\check{\underline{\mu}}}.$$

It is now natural to follow the footsteps of Section 6.3.2 to find the displacement field $\check{\underline{v}} : \overline{S}_z \to \mathbb{R}^{\underline{d}}$ that proves to be the elastic deformation of the grid-shell subjected to the elevated load $\underline{F}_z$. To put it differently we are looking for $\check{\underline{v}}$ for which there exists a stress function $\check{\sigma} \in L^2_{\check{\underline{\mu}}}(\overline{S}_z; \mathcal{S}^{\underline{d}\times\underline{d}}_+)$ such that the equilibrium equation and the constitutive law of elasticity are met:

$$-\mathrm{Div}\big(\check{\underline{\sigma}}\check{\underline{\mu}}\big) = \underline{F}_z \ \text{ in } \ \mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0, \qquad \check{\underline{\sigma}} \in j_+\big(\check{\underline{\mathscr{H}}}, \underline{e}(\check{\underline{v}})\big) \quad \check{\underline{\mu}}\text{-a.e.} \tag{6.54}$$

Of course the Hooke law above requires that the function $\check{\underline{v}}$ is at least C^1. Meanwhile one can expect that the displacement function in question should be again expressed by: $\check{\underline{v}}(\underline{x}) = \frac{\sqrt{2}\mathcal{Z}}{C_0} \big(w(x) + \sqrt{2}\, u(x)\, \mathrm{e}_{\underline{d}} \big)$. This time around, however, we must deal with an arbitrary pair $(u, w) \in \overline{\mathcal{X}}_1$ therefore $\check{\underline{v}}$ may be discontinuous as an element of $BV \cap L^\infty(\underline{\Omega}; \mathbb{R}^{\underline{d}})$. At the present time we lack the tools of calculus that would make it possible to apply the strain operator $\underline{e}$ to such function $\underline{v}$ so that the output is meaningful $\check{\underline{\mu}}$ -a.e., see Section 5.7.2 for a similar discussion.

By taking advantage of the specific truss-like form of $\underline{\lambda}_{\underline{\mathscr{S}}}$ our intention will now be to rewrite the equations of elasticity (6.54) to an equivalent form that can be then relaxed to adhere to a non-smooth $\check{\underline{v}}$ function:

Proposition 6.33. *For a smooth displacement function $\check{\underline{v}} \in \mathcal{D}(\mathbb{R}^{\underline{d}} \setminus \Upsilon_0; \mathbb{R}^{\underline{d}})$ the following conditions are equivalent:*

(i) there exists a stress function $\check{\underline{\sigma}} \in L^2_{\underline{\mu}}(\overline{S}_z; \mathcal{S}^{d\times d}_+)$ such that (6.54) holds true;
(ii) the function $\check{\underline{v}}$ solves the maximization problem in (6.2) for $\underline{\lambda} = \underline{\lambda}_{\underline{\mathscr{S}}}$ and $\underline{F} = \underline{F}_z$;
(iii) for $\underline{\mathscr{S}}$-a.e. pair $(\underline{x}, \underline{y})$ there holds

$$\big\langle \underline{e}(\check{\underline{v}}), \underline{\tau}^{\underline{x},\underline{y}} \otimes \underline{\tau}^{\underline{x},\underline{y}} \big\rangle = \frac{\sqrt{2}\,\mathcal{Z}}{C_0} \qquad \text{on} \quad [\underline{x}, \underline{y}]. \tag{6.55}$$

Proof. The equivalence between (i) and (ii) is a smooth variant of the "moreover part" in Proposition 4.17. It can be proved directly through Proposition 4.4 for $j = j_+$. The function $\underline{v} = \check{\underline{v}}$ satisfies (ii) if and only if we have equalities in the chain (6.53). From the observation below (6.53) this implies (iii). Conversely, if (iii) is true then the second inequality in (6.53) is an equality while the positive part $(\,\cdot\,)_+$ in the second line can be disposed of, which makes it equal to $\int \langle \check{\underline{v}}, \underline{F}_z \rangle - \int j\big(\underline{\lambda}_{\underline{\mathscr{S}}}, \underline{e}(\check{\underline{v}})\big)$ (we use the fact that $\int j\big(\underline{\lambda}_{\underline{\mathscr{S}}}, \underline{e}(\check{\underline{v}})\big) = \frac{1}{2}\langle \underline{e}(\check{\underline{v}}) \otimes \underline{e}(\check{\underline{v}}), \underline{\lambda}_{\underline{\mathscr{S}}} \rangle_{(C,\mathcal{M})}$). Since $j_+(\underline{H}, \,\cdot\,) \le j(\underline{H}, \,\cdot\,)$ the latter expression bounds the first line of (6.53) from below thus rendering the first inequality in (6.53) an equality as well, hence the condition (ii). □

Remark 6.34. The condition (iii) above may be verbalized as follows: *the axial component of the strain is uniform in all the bars of the (generalized) truss structure $\underline{\lambda}_{\underline{\mathscr{S}}}$.* From the point of view of Michell theory (see [118]) this is obvious when the truss consists of a finite number of bars: at optimality the cross-sectional stiffnesses $\frac{C_0}{\sqrt{2}\mathcal{Z}}\underline{\mathscr{S}}(dxdy)$ are proportional to the axial forces $\underline{\mathscr{S}}(dxdy)$, which render the axial relative stresses uniform and the same must concern the strains by the Hooke law.

Let $(u, w) \in \overline{\mathcal{X}}_1$ be the optimal pair of functions that were chosen in Section 6.4.1. We convert w to a multifunction $\mathbf{w} := \mathbf{i}^{-1}(w) \in \mathbf{W}_\Omega$ that in general may be multivalued. We define a multifunction $\mathbf{w} : \overline{\Omega} \to 2^{\mathbb{R}^d}$ that does not depend on the $\underline{d}$-th coordinate:

$$\check{\mathbf{v}}(\underline{x}) = \frac{\sqrt{2}\,\mathcal{Z}}{C_0}\Big(\mathbf{w}(x) + \big\{\sqrt{2}\,u(x)\,\mathrm{e}_{\underline{d}}\big\}\Big), \tag{6.56}$$

namely for any $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$ we put $\check{\underline{\mathbf{v}}}\big((x,t)\big) = \frac{\sqrt{2}\mathcal{Z}}{C_0}\big\{x' + \sqrt{2}\,u(x)\,\mathrm{e}_{\underline{d}} \,:\, x' \in \mathbf{w}(x)\big\}$. The symbol $\check{\underline{\mathbf{v}}}$ should not be confused with a maximal monotone map, cf. the notation in Section 5.3.2. Next, for any pair $(\underline{x}, \underline{y}) \in \overline{\Omega} \times \overline{\Omega}$ we define the univariate multifunction $\check{\underline{\mathbf{v}}}_{\underline{x},\underline{y}} : \mathbb{R} \to 2^{\mathbb{R}}$:

$$\check{\underline{\mathbf{v}}}_{\underline{x},\underline{y}}(t) := \big\langle \check{\underline{\mathbf{v}}}\big(\underline{x} + t\,\underline{\tau}^{\underline{x},\underline{y}}\big), \underline{\tau}^{\underline{x},\underline{y}} \big\rangle.$$

Theorem 6.35. *For $\underline{\mathscr{S}}$-a.e. pair $(\underline{x}, \underline{y}) \in \overline{\Omega} \times \overline{\Omega}$ the following facts hold true:*

(i) The relative elongations of bars are uniform, namely

$$\frac{\big\langle \underline{y}' - \underline{x}', \underline{\tau}^{\underline{x},\underline{y}} \big\rangle}{\big|\underline{y} - \underline{x}\big|} = \frac{\sqrt{2}\,\mathcal{Z}}{C_0} \qquad \forall\, \underline{x}' \in \check{\underline{\mathbf{v}}}(\underline{x}),\ \forall\, \underline{y}' \in \check{\underline{\mathbf{v}}}(\underline{y}).$$

(ii) the map $\check{\underline{\mathbf{v}}}_{\underline{x},\underline{y}}$ is univalued on the interval $\big[0,|\underline{y}-\underline{x}|\big]$. Its unique selection $\underline{v}_{\underline{x},\underline{y}}$ is affine on this interval while

$$\underline{v}'_{\underline{x},\underline{y}} = \frac{\sqrt{2}\,\mathcal{Z}}{C_0} \qquad \text{on} \quad \big[0,|\underline{y}-\underline{x}|\big]. \tag{6.57}$$

Proof. For simplicity we shall assume that $\frac{\sqrt{2}\,\mathcal{Z}}{C_0}=1$ in the proof. Owing to Remark 6.28 we must prove the assertions (i), (ii) for pairs $\big(\underline{x},\underline{y}\big)=\big(\underline{z}(x),\underline{z}(y)\big)$ for $\mathscr{S}$-a.e. $(x,y)\in\overline{\Omega}\times\overline{\Omega}$. For any such pair (x,y) we have $\underline{x}'\in\underline{v}\big(\underline{z}(x)\big)$ if and only if $\underline{x}'=x'+\sqrt{2}\,u(x)\,\mathrm{e}_{\underline{d}}$ and $x'\in\mathbf{w}(x)$; similar conclusion holds for y. As a result we may compute:

$$\begin{aligned}
\big\langle \underline{y}'-\underline{x}', \underline{z}(y)-\underline{z}(x)\big\rangle &= \langle y'-x', y-x\rangle + \sqrt{2}\,\big(u(y)-u(x)\big)\big(z(y)-z(x)\big)\\
&= \langle y'-x', y-x\rangle + \big(u(y)-u(x)\big)^2\\
&= \left(\frac{1}{2}\big(u(y)-u(x)\big)^2 + \langle y'-x', y-x\rangle\right) + \frac{1}{2}\big(u(y)-u(x)\big)^2\\
&= |y-x|^2 + \big(z(y)-z(x)\big)^2 = |\underline{z}(y)-\underline{z}(x)|^2.
\end{aligned}$$

To pass to the last line we used Theorem 5.95, but in the format discussed in Remark 5.96: particularly we exploited the condition (iii', v'). The assertion (i) here follows by comparing the left hand side and the right hand side of the chain above and then by dividing by $|\underline{z}(y)-\underline{z}(x)|^2$.

The fact that $\check{\underline{\mathbf{v}}}_{\underline{x},\underline{y}}$ is univalued and affine on $\big[0,|\underline{y}-\underline{x}|\big]$ is straightforward due to statements (2), (3) in Theorem 5.95. Then, the constant value $\underline{v}'_{\underline{x},\underline{y}}$ must be equal to $\big\langle \underline{y}'-\underline{x}', \underline{\tau}^{\underline{x},\underline{y}}\big\rangle/|\underline{y}-\underline{x}|$ for any $\underline{x}'\in\check{\underline{\mathbf{v}}}(\underline{x})$ and $\underline{y}'\in\check{\underline{\mathbf{v}}}(\underline{y})$. The latter value is equal to 1 owing to the claim (i). □

It is straightforward to see that when u,w are of class C^1 then the multifunction is univalued, i.e. $\check{\underline{\mathbf{v}}}(\underline{x})=\{\check{\underline{v}}(\underline{x})\}$ where $\check{\underline{v}}\in C^1(\overline{\underline{\Omega}};\mathbb{R}^{\underline{d}})$. Then the condition (6.57) is exactly (6.55). We have therefore showed that, although the function $\check{\underline{v}}$ may not be $\check{\underline{\mu}}$-a.e. differentiable in any meaningful sense, the tangential strain along $\mathscr{S}$-a.e. bar $[\underline{x},\underline{y}]$ may be defined and is in fact uniform as required by condition (iii) in Proposition 6.33. On another note, one can show by using Theorem 4.6 that $\check{\underline{\sigma}}$ spoken of in (i) of Proposition 6.33 can be computed as $\check{\underline{\sigma}}=\frac{d\underline{\varsigma}_{\mathscr{S}}}{d\check{\underline{\mu}}}$.

6.4.3. Example of an optimal grid-shell

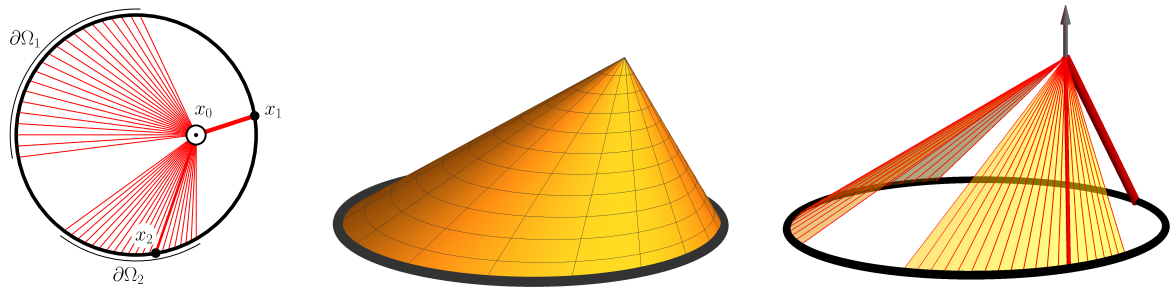
To mirror the structure of Section 6.3 we shall now give an example of an optimal grid-shell constructed from an analytically computed optimal string system. The only such solution has been so far found for the one-point load $f=\delta_{x_0}$, see Section 5.5.3. We will focus on the most interesting example where Ω is a disk. The Prager problem for such data Ω, f is of particular significance as it was the one studied in the work [146] that has started the research on the so called *Prager structures* in the first place. More reference to the analysis from the latter work can be found in Remark 6.37 below.

Example 6.36 (**Optimal form for a one-point load**)**.** We consider the optimal form-finding problem for the data: $\Omega=\big\{|x|<R\,:\,x\in\mathbb{R}^2\big\}$, $f=\delta_{x_0}$ where $x_0\in\Omega$ and

$\underline{\Omega} = \Omega \times (-h_0, h_0)$ where $h_0 = R$. Below we gather the results from Section 5.5.3, starting with a solution (u, w) of the problem $(\overline{\mathscr{P}})$ upon which the elevation function z is built:

$$(u, w) = \Big(\sqrt{2}\, d_0, 2\, x_0\Big)\, h, \qquad z = \frac{1}{\sqrt{2}}\, u, \qquad z(x_0) = d_0 = \sqrt{R^2 - |x_0|^2}$$

where $h : \overline{\Omega} \to \mathbb{R}$ is the function whose graph is a cone that contains the hoop $\underline{\Upsilon}_0 = \partial\Omega \times \{0\}$ and whose vertex is $(x_0, 1)$. The conic graph $\overline{S}_z$ is shown in Figure 6.3(b). Its vertex $\underline{z}(x_0) = (x_0, d_0)$, called *the apical point* of the constructed form, satisfies an interesting geometric property: it is the vertical projection of x_0 onto the top half of the sphere $\big\{\underline{x} \in \mathbb{R}^3 \,:\, |\underline{x}| = R\big\}$.



(a) (b) (c)

Figure 6.3: The form-finding problem for a one-point force $f = \delta_{x_0}$ over a disk: (a) a 2D pre-stress $\varsigma_{\mathscr{S}}$ composed from a string system that $\mathscr{S}$ solves the problem $(\widetilde{\mathscr{P}}^*)$; (b) the graph $\overline{S}_z$ on which the optimal form concentrates; (c) the optimal form – a 3D truss $\underline{\mathscr{S}}$ obtained by elevating the strings $\mathscr{S}$.

For any probability measure $p \in \mathrm{P}(\partial\Omega)$ that satisfies the barycentre condition $x_0 = \int_{\partial\Omega} x\, p(dx)$ the string system $(\pi, \mathscr{S}) = (\alpha\mathscr{S}, \mathscr{S})$ satisfying

$$\mathscr{S} = \frac{|x_0 - x|}{\sqrt{2}\, d_0}\, p(dx) \otimes \delta_{x_0}(dy), \qquad \alpha(x, x_0) = \frac{\sqrt{2}\, d_0}{|x_0 - x|} = \sqrt{2}\, \frac{z(x_0) - z(x)}{|x_0 - x|} \quad \text{for } x \in \partial\Omega.$$

solves the problem $(\widetilde{\mathscr{P}}^*)$; in Fig. 6.3(a) the pre-stress $\varsigma_{\mathscr{S}}$ is showed for a particular p already considered in Section 5.5.3. Above we have acknowledged the fact that $z = 0$ on $\partial\Omega$. We can see that $\chi^*_{\mathrm{K}}\big(\alpha(x, x_0), 1\big) = 1 + \big(\frac{z(x_0) - z(x)}{|x_0 - x|}\big)^2$ which approves of the relation (6.47). The push-forward definition of $\underline{\mathscr{S}}$ in (6.46) yields a simple formula for the 3D truss: it is composed of bars connecting the apical point $\underline{z}(x_0)$, on which the one-point load $\underline{F}_z$ is concentrated, to the hoop $\underline{\Upsilon}_0$, namely

$$\underline{\mathscr{S}} = \frac{|\underline{z}(x_0) - \underline{x}|}{d_0}\, \underline{p}(d\underline{x}) \otimes \delta_{\underline{z}(x_0)}(d\underline{y}), \qquad \underline{F}_z = \mathrm{e}_3\, \delta_{\underline{z}(x_0)}$$

where the probability $\underline{p} \in \mathrm{P}(\underline{\Upsilon}_0)$ is the push-forward of p by $\underline{z}$ (note that $\underline{z}(x) = (x, 0)$ on $\partial\Omega$). The optimal 3D stress field $\underline{\varsigma}_{\underline{\mathscr{S}}}$ representing the optimal form is visualized in Fig. 6.3(c). Clearly all the 3D bars lie on the surface $\overline{S}_z$. By looking at the two 3D bars $[(x_0, d_0), (x_1, 0)]$ and $[(x_0, d_0), (x_2, 0)]$ in Fig. 6.3(c) and comparing their cross-sectional areas to the ones of 2D strings $[x_0, x_1]$ and $[x_0, x_2]$, respectively, one may observe the difference in rescaling that is in agreement with Remark 6.29.

Finally, for each point $\underline{x} \in \overline{S}_z$ let us give the elastic 3D displacement of an elastic form $\underline{\lambda}_{\underline{\mathscr{S}}}$. According to the formula (6.56) (all the multifunctions are univalued here)

$$\check{\underline{v}}(\underline{x}) = \frac{\sqrt{2}\,\mathcal{Z}}{C_0}\, h\big(\underline{z}^{-1}(\underline{x})\big)\,\Big(2x_0 + 2d_0\,\mathrm{e}_3\Big) = \frac{\sqrt{2}\,\mathcal{Z}}{C_0}\, d(\underline{x},\mathbb{R}^2)\,\Big(2x_0 + 2d_0\,\mathrm{e}_3\Big)$$

where $d(\underline{x},\mathbb{R}^2)$ is the height of a point $\underline{x}$. To approve of the strain analysis from Section 6.4.2 we shall geometrically show that each bar $[\underline{z}(x),\underline{z}(x_0)]$, $x \in \partial\Omega$ undergoes the relative axial elongation equal to $\frac{\sqrt{2}\,\mathcal{Z}}{C_0}$ as stated in assertion (i) of Theorem 6.35:

$$\begin{aligned}
&\tfrac{C_0}{\sqrt{2}\,\mathcal{Z}}\Big\langle \check{\underline{v}}\big(\underline{z}(x_0)\big) - \check{\underline{v}}\big(\underline{z}(x)\big)\,,\, \underline{z}(x_0) - \underline{z}(x)\Big\rangle = \Big\langle 2x_0 + 2d_0\,\mathrm{e}_3\,,\,(x_0 - x) + d_0\,\mathrm{e}_3\Big\rangle\\
&= \Big\langle \big(x_0 - x + d_0\,\mathrm{e}_3\big) + \big(x_0 + x + d_0\,\mathrm{e}_3\big)\,,\,(x_0 - x) + d_0\,\mathrm{e}_3\Big\rangle\\
&= |x_0 - x + d_0\,\mathrm{e}_3|^2 + \Big(|x_0|^2 - |x|^2 + (d_0)^2\Big) = |\underline{z}(x_0) - \underline{z}(x)|^2 + \Big(|x_0|^2 - |x|^2 + R^2 - |x_0|^2\Big)
\end{aligned}$$

where the last term vanishes since $|x| = R$ for each $x \in \partial\Omega$. △

Remark 6.37. In case when $\underline{\mathscr{S}}$ is a classical truss, i.e. $\underline{\mathscr{S}}$ is finitely supported, the structure visualized in Fig. 6.3(c) was conjectured to solve the Prager problem already in the paper [146]. The authors of the latter work, however, showed that it is optimal within a fairly narrow class of structures: trusses that are composed of bars connecting the apical point $(x_0,\tilde{h})$ to points on the supporting hoop Υ_0 (with the elevation parameter $\tilde{h}$ being a design variable). They did not succeed in proving the so called "global optimality". Their conjecture is confirmed here for the first time. One should bear in mind, however, that apart from the construction above a considerable effort was made to show that the pairs $(u,w),(\pi,\mathscr{S})$ solve $(\overline{\mathscr{P}}),(\widetilde{\mathscr{P}^*})$, respectively, see Section 5.5.3.

6.5. Numerical approximation of optimal forms by grid-shells

Let us now address the matter of choosing the numerical strategy for solving the form-finding problems (PFF), (EFF); we will concentrate on the 3D setting, i.e. when $d = 2$ and $\underline{d} = 3$. Upfront one must realize that any formulation which involves looking for a surface S_z (constituting a membrane shell) entails great difficulties e.g. due to the non-convexity, see [23], [130]. It is thus reasonable to numerically tackle the relaxed problems instead, i.e. the Prager problems (PPP), (EPP). Up to the anisotropy setting (usually isotropic material is assumed) the latter approach is extensively exploited in the literature, also for $d = 1$, $\underline{d} = 2$. The three dimensional space is discretized using either finite elements (see [88], [56]) or a 3D ground structure (see [92], [107], [117], [122]). In those works the 3D bodies thus numerically generated tend to oscillate about a single 2D surface. Unavoidably the burden of 3D discretization affects the computational efficiency – the mesh/ground structure resolutions are usually unsatisfactory, especially considering that we wish to extract a 2D form.

In this chapter the aforementioned "dimensionality issues" are being circumvented. The 3D form-finding problems (PFF), (EFF) together with the 3D Prager problems (PPP), (EPP) have been reduced to the 2D optimal membrane problem in the form of

convex pair $(\overline{\mathscr{P}}), (\mathscr{P}^*)$ (or $(\overline{\mathscr{P}}), (\widetilde{\mathscr{P}^*})$ where an optimal string system is searched for) heavily studied in Chapter 5. We can thus concentrate on numerical methods for the latter 2D problems. This has been already achieved in Section 5.5.4 where the ground structure method was applied: in the $(\overline{\mathscr{P}}_h), (\widetilde{\mathscr{P}^*_h})$ we seek an optimal string system assuming that its joints are in the finite set X^h. In Section 6.4 we have shown that a string system solving $(\widetilde{\mathscr{P}^*})$ generates an optimal form/grid-shell if suitably elevated. If we repeat the elevation procedure for the system solving $(\widetilde{\mathscr{P}^*_h})$ we may expect to arrive at a grid-shell approximation of a form that solves the problems (PFF), (EFF) or (PPP), (EPP).

6.5.1. From a suboptimal 2D string system to a suboptimal 3D grid-shell

For a convex bounded domain $\Omega \subset \mathbb{R}^2$ and a load $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ we are to construct a grid-shell approximations of solutions of the Prager problems (PPP), (EPP) for a 3D domain $\underline{\Omega} = \Omega \times (-h_0, h_0)$.

Let $\mathrm{X}^h \subset \mathbb{R}^2$ (the planar nodal grid) be a finite subset of $\overline{\Omega}$ that satisfies the condition (5.106) while $f_h \in \mathcal{M}(\mathrm{X}^h; \mathbb{R})$ will be the discretization of the load f. Further, to avoid repetitions, we agree that all the objects and notations from Section 5.5.4 are adopted here. Accordingly let us choose and fix pairs

$$(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2) \in \mathbb{R}^{3\times m} \quad \text{and} \quad (\mathbf{r}, \mathbf{s}, \mathbf{p}) \in \mathbb{R}^N_+ \times \mathbb{R}^N \times \mathbb{R}^N_+ \qquad \text{that solve} \qquad \boldsymbol{P}_h \ \text{ and } \ \boldsymbol{P}^*_h,$$

respectively. By Theorem 5.66 such solutions always exist and they satisfy the system of optimality conditions in (5.113). Let us stress that the construction to be put forth below will not rely on any additional assumptions (contrarily to the constructions in Section 6.3 and 6.4).

The construction of a grid-shell approximation of an optimal form starts below. While we go through the process we will be referring to a particular example where Ω is a square and $f = \delta_{x_0} + \mathcal{H}^1 \llcorner [a, b]$ for $x_0 \in \Omega$ and $a, b \in \partial\Omega$. The suboptimal pre-stress $\varsigma_{\mathscr{S}_h}$ was already showed in Fig. 5.12(b) (for the mesh parameter $h = 1/100$) and is now repeated in Fig 6.4(b) for a higher resolution: $h = 1/200$. Note that the part $\mathcal{H}^1 \llcorner [a, b]$ of the load is discretized to multiple point loads, which is visualized by small dots in Fig. 6.4(b).

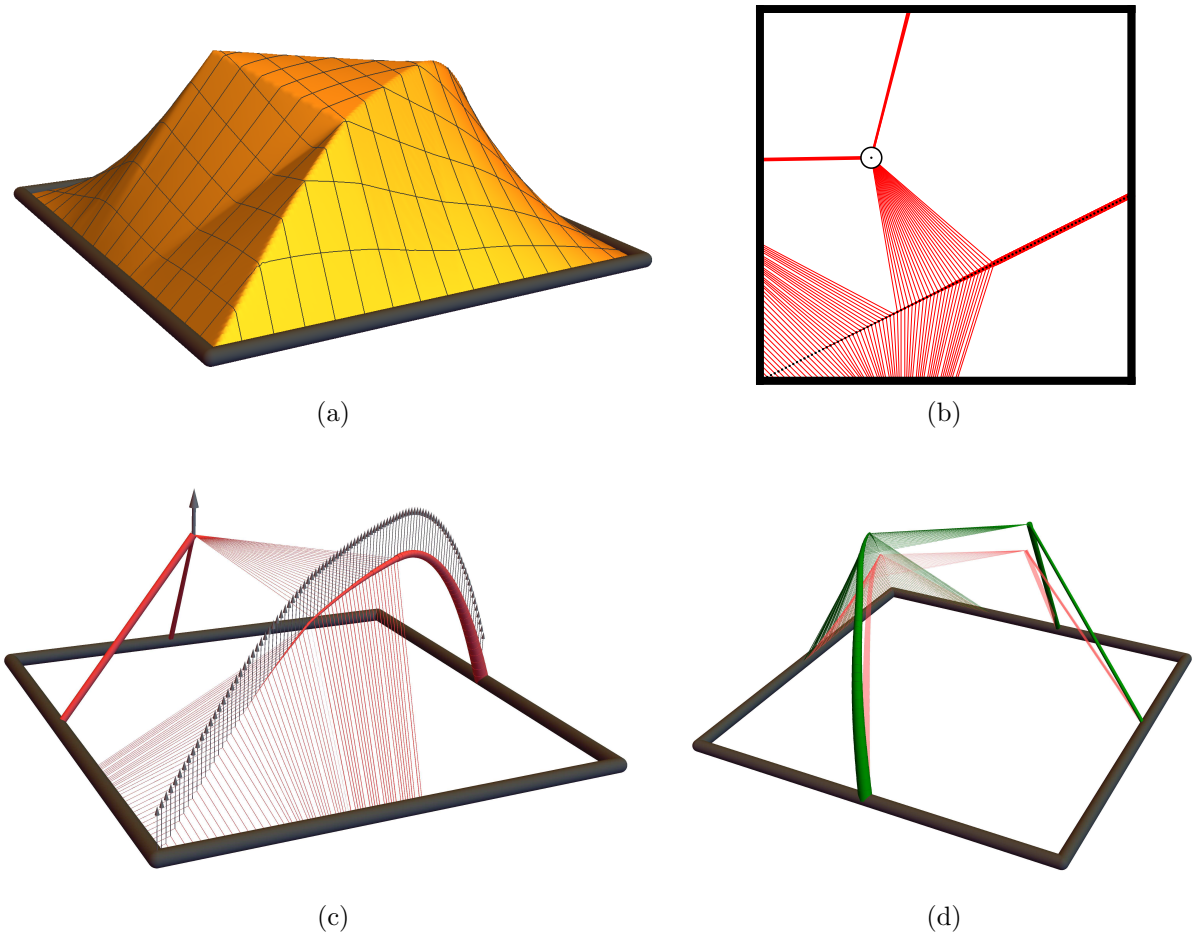
The construction of the elevation function known from the previous sections is now brought to the discrete setting, i.e.

$$\mathbf{z} := \frac{1}{\sqrt{2}}\,\mathbf{u}, \qquad z_h := \iota^h_\perp(\mathbf{z}), \qquad \underline{z}_h(x) := \big(x, z_h(x)\big)$$

where $\iota^h_\perp$ is the interpolation operator which produces a Lipschitz continuous function that is piece-wise affine (see Section 5.5.4). The graph $\overline{S}_{z_h}$ of the interpolated function z_h is showed in Fig. 6.4(a). It is the target surface on which the grid-shell to be built *should* concentrate on.

Let us recall that each bar in the planar ground structure spanning X^h is identified by an index $k \in \{1, \dots, N\}$ while $\big(\mathrm{x}(k), \mathrm{y}(k)\big) \in \overline{\Omega} \times \overline{\Omega}$ are the end-points of this bar. To each such planar bar there will now correspond an elevated bar k whose end-points are computed as follows:

$$\underline{\mathrm{x}}_h(k) := \underline{z}_h\big(\mathrm{x}(k)\big), \qquad \underline{\mathrm{y}}_h(k) := \underline{z}_h\big(\mathrm{y}(k)\big).$$



(a) (b)

(c) (d)

Figure 6.4: Construction of a grid-shell approximating an optimal form for the load data $f = \delta_{x_0} + \mathcal{H}^1 \llcorner [a,b]$: (a) the surface $\overline{S}_{z_h}$; (b) the pre-stress $\varsigma_{\mathscr{S}_h}$ approximating a solution of $(\mathscr{P}^*)$; (c) the plastic grid-shell $\underline{\varsigma}_{\mathscr{S}_h}$ obtained via elevation of the string system $\mathscr{S}_h$ and the elevated load $\underline{F}_h$; (d) deformation $\check{\underline{v}}_h$ of the elastic grid-shell $\underline{\lambda}_{\mathscr{S}_h}$ (different perspective).

Next we define the elevated discrete load

$$\underline{F}_h := \mathrm{e}_3\, \underline{z}_h \,\#\, f_h = \sum_{j=1}^{m} \mathrm{e}_3\, \mathbf{f}(j)\, \delta_{\underline{z}_h(v_j)},$$

cf. Section 5.5.4 for definitions of $\mathbf{f}$ and v_j. It is clear that $\underline{F}_h \in \mathscr{T}(\Omega, f_h)$. The finite planar string system showed in Fig. 6.4(b) may be computed as $\mathscr{S}_h = \sum_{k=1}^{N} \mathbf{s}(k)\, \delta_{(\mathrm{x}(k),\mathrm{y}(k))}$. To produce the 3D elevated truss $\underline{\mathscr{S}}_h$ we must follow the push-forward formula in (6.46), namely

$$\underline{\mathscr{S}}_h = (\underline{z}_h, \underline{z}_h) \,\#\, \Big(\sqrt{2}\, \big(\chi^*_{\mathrm{K}}(\alpha_h, 1)\big)^{\frac{1}{2}}\, \mathscr{S}_h \Big) = \sum_{k=1}^{N} \underline{\mathbf{s}}(k)\, \delta_{(\underline{\mathrm{x}}_h(k), \underline{\mathrm{y}}_h(k))}$$

where owing to the optimality condition (iv) in (5.113) we can put $\alpha_h\big(\mathrm{x}(k), \mathrm{y}(k)\big) = (\mathbf{Hu})(k)/\mathbf{L}(k)$ for every k. As a result we arrive at the formula on the right hand side above where

$$\underline{\mathbf{s}}(k) = \sqrt{2} \sqrt{1 + \left(\frac{(\mathbf{Hz})(k)}{\mathbf{L}(k)} \right)^2}\, \mathbf{s}(k) = \sqrt{\big(\sqrt{2}\, \mathbf{s}(k)\big)^2 + \big(\mathbf{q}(k)\big)^2}$$

and the latter Pythagorean formula is again due to (iv) in (5.113).

The construction process is complete. We may now compute the stress and Hooke fields $\underline{\varsigma}_h := \underline{\varsigma}_{\underline{\mathscr{S}}_h}$ and, respectively, $\underline{\lambda}_h := \underline{\lambda}_{\underline{\mathscr{S}}_h}$ following the integration formulas (6.49) and (6.52). Due to the finite supports of the pairing measure we arrive at

$$\underline{\varsigma}_{\underline{\mathscr{S}}_h} = \sum_{k=1}^{N} \underline{\mathbf{s}}(k)\, \underline{\varsigma}^{\underline{\mathrm{x}}_h(k),\, \underline{\mathrm{y}}_h(k)}, \qquad \underline{\lambda}_{\underline{\mathscr{S}}_h} = \frac{C_0}{\sqrt{2}\,\mathcal{Z}_h} \sum_{k=1}^{N} \underline{\mathbf{s}}(k)\, \underline{\lambda}^{\underline{\mathrm{x}}_h(k),\, \underline{\mathrm{y}}_h(k)}$$

where $\mathcal{Z}_h = \max \boldsymbol{P}_h = \min \boldsymbol{P}_h^*$. The grid-shell $\underline{\varsigma}_{\underline{\mathscr{S}}_h}$ (plastic) or $\underline{\lambda}_{\underline{\mathscr{S}}_h}$ (elastic) are visualized in Fig. 6.4(c) together with the elevated discrete load $\underline{F}_h$.

Finally, the displacement function $\check{\underline{v}}_h \in \mathrm{Lip}(\underline{\Omega}; \mathbb{R}^3)$ can be computed as

$$\check{\underline{v}}_h(\underline{x}) := \frac{\sqrt{2}\,\mathcal{Z}_h}{C_0}\Big(w_h(x) + \sqrt{2}\, u_h(x)\, \mathrm{e}_3\Big), \qquad w_h := \iota_{\shortparallel}^h(\mathbf{w}_1, \mathbf{w}_2), \quad u_h := \iota_{\perp}^h(\mathbf{u}),$$

namely $\check{\underline{v}}_h$ does not depend on the third coordinate. Considering the optimality condition (iii) in (5.113) we can repeat the proof of the assertion (i) in Theorem 6.35 to find that the relative elongations of the bars in the grid-shell are uniform:

$$\frac{\Big\langle \check{\underline{v}}\big(\underline{\mathrm{y}}_h(k)\big) - \check{\underline{v}}\big(\underline{\mathrm{x}}_h(k)\big)\,,\, \underline{\tau}^{\underline{\mathrm{x}}_h(k),\underline{\mathrm{y}}_h(k)} \Big\rangle}{\big|\underline{\mathrm{y}}_h(k) - \underline{\mathrm{x}}_h(k)\big|} = \frac{\sqrt{2}\,\mathcal{Z}_h}{C_0} \qquad \forall\, k \text{ such that } \underline{\mathbf{s}}(k) \neq 0.$$

The elastic deformation $\check{\underline{v}}_h$ is visualized in Fig. 6.4(d): the dark green colour shows the deformed state.

By repeating the arguments from Section 6.4 we show that the constructed grid-shell approximations are admissible for the Prager problems if we consider the discretized load f_h instead of the original one:

Proposition 6.38. *For a suitably large height parameter h_0 the pairs $(\underline{\varsigma}_{\underline{\mathscr{S}}_h}, \underline{F}_h)$ and $(\underline{\lambda}_{\underline{\mathscr{S}}_h}, \underline{F}_h)$ are admissible for the Prager problems* (PPP) *and, respectively,* (EPP) *for data Ω, f_h, i.e.*

$$\underline{F}_h \in \mathscr{T}(\underline{\Omega}, f_h), \qquad -\mathrm{Div}\, \underline{\varsigma}_{\underline{\mathscr{S}}_h} = \underline{F}_h \quad \text{in } \mathbb{R}^3 \setminus \underline{\Upsilon}_0, \qquad \int_{\underline{\Omega}} \mathrm{Tr}\, \underline{\lambda}_{\underline{\mathscr{S}}_h} = C_0.$$

Moreover

$$\int_{\underline{\Omega}} \mathrm{Tr}\, \underline{\varsigma}_{\underline{\mathscr{S}}_h} = \sqrt{2}\,\mathcal{Z}_h, \qquad \underline{\mathcal{C}}(\underline{\lambda}_{\underline{\mathscr{S}}_h}, \underline{F}_h) \leq \frac{(\mathcal{Z}_h)^2}{C_0}.$$

Proof. One may take $h_0 = \|z_h\|_{\infty,\Omega}$. The fact that $\underline{F}_h \in \mathscr{T}(\underline{\Omega}, f_h)$ follows directly from (6.4). With f being replaced by f_h the pairs (u_h, w_h) and $(\pi_h, \mathscr{S}_h)$ solve the problems $(\overline{\mathscr{P}}_h)$ and $(\widetilde{\mathscr{P}}_h^*)$, respectively, in particular $(\pi_h, \mathscr{S}_h) \in \Xi_{\mathrm{X}^h \times \mathrm{X}^h}(\Omega, f_h) \subset \Xi(\Omega, f_h)$. Apart from (ii) all the optimality conditions in (5.99) are met and only these were used to prove the first part of Lemma 6.30 hence in the same way we prove that $-\mathrm{Div}\, \underline{\varsigma}_{\underline{\mathscr{S}}_h} = \underline{F}_h$ in $\mathbb{R}^3 \setminus \underline{\Upsilon}_0$. Similarly we use the arguments from the proof of the lemma to show that $\int_{\underline{\Omega}\times\underline{\Omega}} |\underline{y} - \underline{x}|\, \mathscr{S}_h(d\underline{x}d\underline{y}) = \sqrt{2}\, \min \widetilde{\mathscr{P}}_h^* = \sqrt{2}\,\mathcal{Z}_h$ and thus automatically $\int_{\underline{\Omega}} \mathrm{Tr}\, \underline{\varsigma}_{\underline{\mathscr{S}}_h} = \sqrt{2}\,\mathcal{Z}_h$ and $\int_{\underline{\Omega}} \mathrm{Tr}\, \underline{\lambda}_{\underline{\mathscr{S}}_h} = C_0$. The chains of inequalities in the proof of Theorem 6.32 remain valid if we replace $\mathscr{S}, \underline{F}_z, \mathcal{Z}$ by $\mathscr{S}_h, \underline{F}_h, \mathcal{Z}_h$ hence the inequality $\underline{\mathcal{C}}(\underline{\lambda}_{\underline{\mathscr{S}}_h}, \underline{F}_h) \leq \frac{(\mathcal{Z}_h)^2}{C_0}$. □

Similarly as in the case of any other numerical method developed in this thesis the convergence of this technique will not be studied either, although we can make predictions. It is a fact that $f_h \overset{*}{\rightharpoonup} f$ therefore (possibly after extracting a subsequence) the load $\underline{F}_h$ must weakly-* converge to a load $\underline{F} \in \mathscr{T}(\Omega, f)$. Since we may expect that $\mathcal{Z}_h$ approaches $\mathcal{Z}$ when $h \to 0$ (see Conjecture 5.68) then based on the foregoing proposition we can hope that the construction of grid-shells provides an efficient numerical method for (PPP), (EPP), namely that $\underline{\varsigma}_{\mathscr{S}_h}, \underline{\lambda}_{\mathscr{S}_h}$ are suboptimal for the respective Prager problems. The situation becomes more delicate if we are looking at the designs $\underline{\varsigma}_{\mathscr{S}_h}, \underline{\lambda}_{\mathscr{S}_h}$ from perspective of the form-finding problems (PFF), (EFF), namely

$$\text{there is no guarantee that:} \qquad \operatorname{spt}\underline{\varsigma}_{\mathscr{S}_h} \subset \overline{S}_{z_h}, \qquad \operatorname{spt}\underline{\lambda}_{\mathscr{S}_h} \subset \overline{S}_{z_h}$$

or, equivalently, it is not clear if for all k such that $\mathbf{s}(k) \neq 0$ the segments $\big[\underline{z}\big(\mathrm{x}(k)\big), \underline{z}\big(\mathrm{y}(k)\big)\big]$ are contained in the surface $\overline{S}_{z_h}$. Therefore the structures $\underline{\varsigma}_{\mathscr{S}_h}, \underline{\lambda}_{\mathscr{S}_h}$ might not be *forms* in the understanding of Section 6.1. By using the optimality conditions in (5.99) one can show that if a point $x \in [\mathrm{x}(k), \mathrm{y}(k)]$ belongs to the grid X^h then there must hold $\underline{z}(x) \in \big[\underline{z}\big(\mathrm{x}(k)\big), \underline{z}\big(\mathrm{y}(k)\big)\big]$. If, however, $x \notin \mathrm{X}^h$ then there is no way of ensuring the same, since the piece-wise affine interpolation $z_h = \iota^h_\perp(\mathbf{z})$ is chosen heuristically. On the other hand, for small parameters h there is always a point $\tilde{x} \in \mathrm{X}^h$ which lies in close vicinity of x. It is therefore justified to hope that the number $\sup_{\underline{x} \in \operatorname{spt}\underline{\varsigma}_h} d(\underline{x}, \overline{S}_{z_h})$ tends to zero when $h \to 0$. This is in fact what is observed in the numerical simulations (see Fig. 6.5 below) – for small h the deviation between truss as in Fig. 6.4(c) and graph as in Fig. 6.4(a) is impossible to discern visually. Those experiments encourage to predict the following:

Conjecture 6.39. *For X^h being a family of nodal grids that satisfy condition* (5.106) *let $z_h, \mathscr{S}_h, \underline{F}_h$ be constructed as above. Then there exist: $z \in C_0(\Omega)$, $\underline{\varsigma} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{3\times 3}_+)$, $\underline{\lambda} \in \mathcal{M}(\overline{\Omega}; \mathcal{K}_{fib})$ such that:*

(i) $(\underline{\varsigma}, \underline{F}_z)$ and $(\underline{\lambda}, \underline{F}_z)$ solve the Prager problems (PPP) *and, respectively,* (EPP) *where $\underline{F}_z = \mathrm{e}_{\underline{d}}\, \underline{z} \,\#\, f$;*

(ii) $\underline{\varsigma}$ and $\underline{\lambda}$ are forms concentrated on $\overline{S}_z$, namely

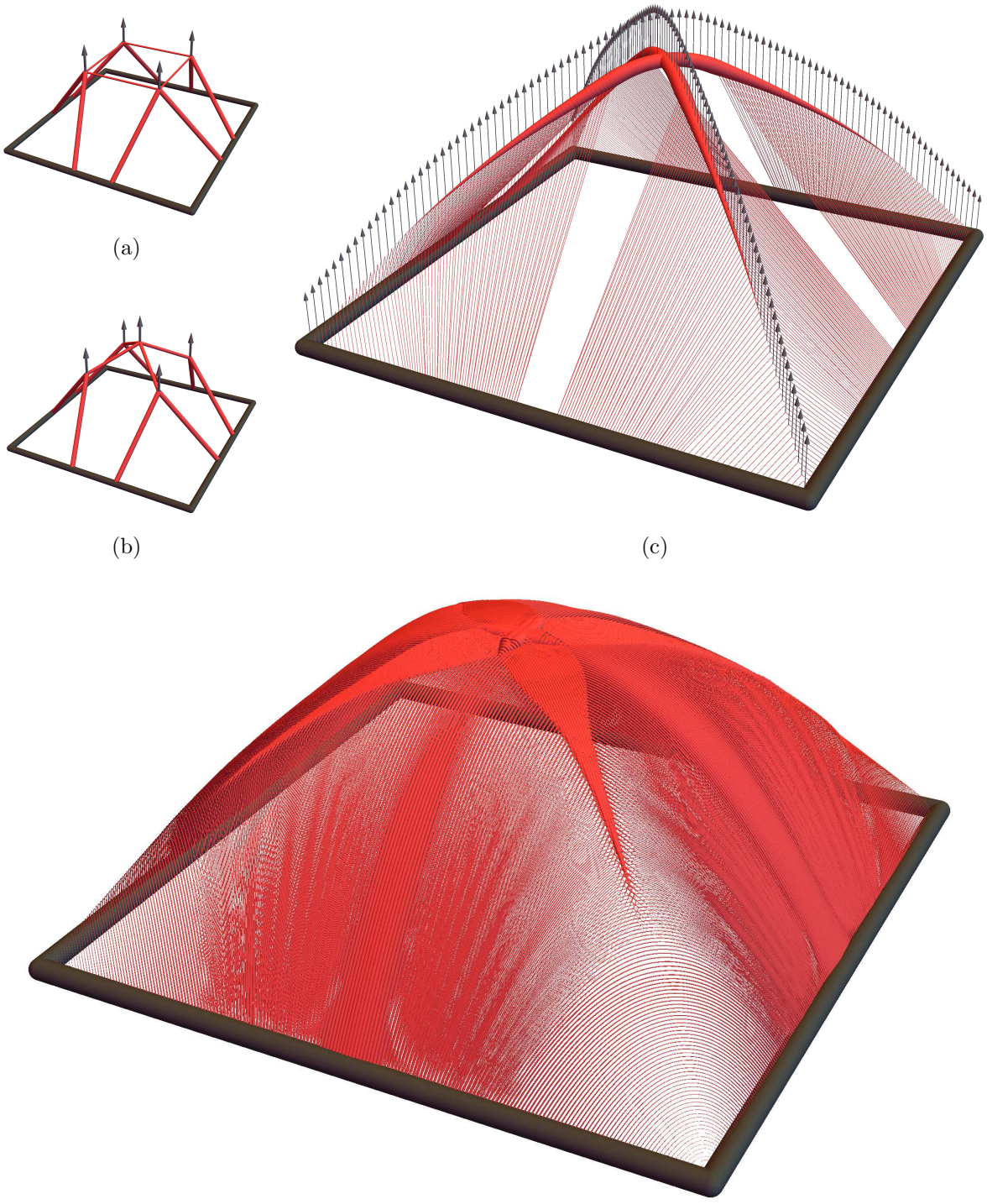
$$\operatorname{spt}\underline{\varsigma} \subset \overline{S}_z \qquad \text{and} \qquad \operatorname{spt}\underline{\lambda} \subset \overline{S}_z;$$

(iii) for h approaching zero (possibly upon choosing subsequence)

$$\underline{\varsigma}_{\mathscr{S}_h} \overset{*}{\rightharpoonup} \underline{\varsigma}, \qquad \underline{\lambda}_{\mathscr{S}_h} \overset{*}{\rightharpoonup} \underline{\lambda}, \qquad \underline{F}_h \overset{*}{\rightharpoonup} \underline{F}_z,$$
$$\overline{S}_{z_h} \to \overline{S}_z \quad \textit{in the sense of Kuratowski.}$$

6.5.2. An overview of grid-shells approximating the optimal forms

We have established that the numerical method for the 3D form-finding problems or the 3D Prager problems reduces to the numerical method that has been already developed for the 2D optimal membrane problem. The additional steps that have to be taken in order to pass from an suboptimal planar string system to a suboptimal spatial grid-shell – described in the previous subsection – are purely algebraic. It is therefore natural to



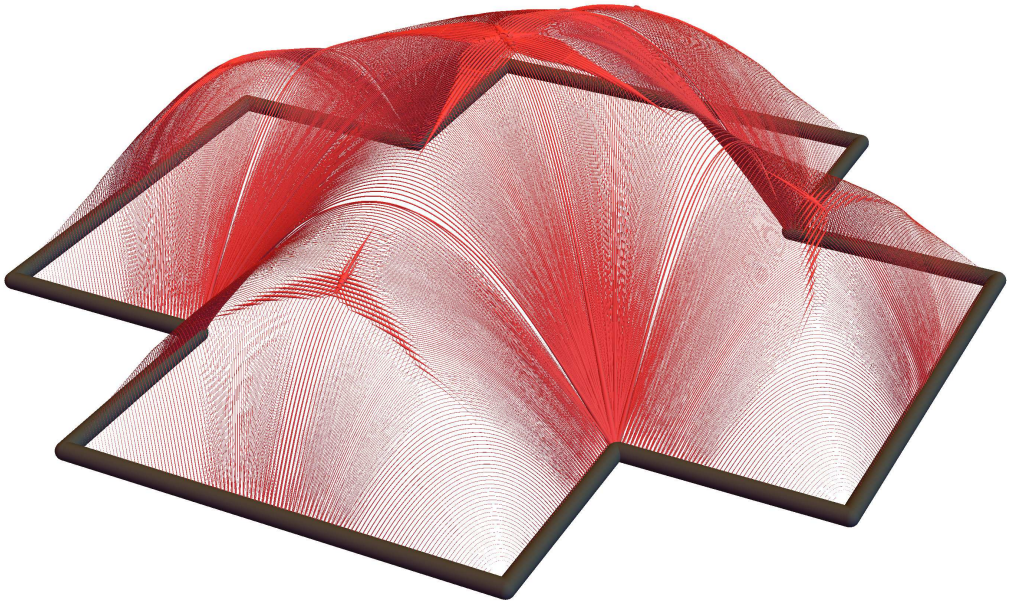
(a)

(b)

(c)

(d)

Figure 6.5: Grid-shell approximations of optimal forms: (a) square domain Ω and a four-force load $f = \sum_{i=1}^{4} \delta_{x_i}$ (Example 5.70); (b) square domain Ω and a five-force load $f = \delta_{x_0} + \sum_{i=1}^{4} \delta_{x_i}$ (Example 5.71); square domain Ω and the diagonal load $f = \mathcal{H}^1 \llcorner \big([a_1, a_3] \cup [a_2, a_4]\big)$ (Example 5.72); (d) square domain Ω and a uniform load $f = \mathcal{L}^2 \llcorner \Omega$ (Example 5.73).



(e)



(f)

Figure 6.5: (cont.) Grid-shell approximations of optimal forms: (e) cross-shaped domain Ω and a uniform load $f = \mathcal{L}^2 \llcorner \Omega$ (Example 5.75); (f) square domain Ω, a uniform load $f = \mathcal{L}^2 \llcorner \Omega$ and an eight point Dirichlet region $\underline{\Upsilon}_0$ (Example 5.117).

revisit the already found numerical solutions of optimal membranes from Section 5.5.6 and perform the elevation procedure to arrive at the grid-shells approximating the optimal forms for the corresponding data Ω, f.

There is no need to repeat the Table 5.1 on computational efficiency – from the additional process the most time consuming is the 3D rendering that was performed in Mathematica® 11.1. To enhance the 3D visual effects we shall employ the results $(\mathbf{u}, \mathbf{w}_1, \mathbf{w}_2)$, $(\mathbf{r}, \mathbf{s}, \mathbf{p})$ computed for the highest resolution, i.e. for $h = 1/200$ (note that the 2D figures in Section 5.5.6 were generated for the intermediate resolution $h = 1/100$). All the grid-shells approximating optimal forms are showed in Fig. 6.5. Each visualization corresponds to a particular example from Section 5.5.6, see the caption of Fig. 6.5 for the precise references. In those examples one may find figures of the suboptimal pre-stress fields $\varsigma_{\mathscr{S}_h}$ and graphs of functions u_h, note that the target surfaces $\overline{S}_{z_h}$ are shallower by a factor of $\sqrt{2}$.

The last grid-shell herein displayed, see Fig. 6.5(f), is to approximate a solution of a Prager problem for the Dirichlet region $\Upsilon_0 = \{\underline{x}_i\}_{i=1}^8 = \{(x_i, 0)\}_{i=1}^8$ where $x_i \in \partial\Omega$, i.e. the structure to be found may rest on eight points. Naturally, such a modification of the Prager problem (where $\Upsilon_0 \neq \partial\Omega \times \{0\}$) requires additional study, e.g. in terms of existence of solution. This is a similar issue to the one considered in Section 5.7.3, where in fact the numerical solution was found for the membrane problem that is pinned at $\Upsilon_0 = \{x_i\}_{i=1}^8$ exactly. With such a numerical solution available the construction of the suboptimal grid-shell remains the same, hence the picture in Fig. 6.5(f) was possible to produce. The theoretical aspects of such alteration of the form-finding problem will be not dwelt on here.

The remarks made in Section 5.5.6 about the examples of suboptimal string systems can be pretty much rewritten for suboptimal grid-shells herein presented. Above all, the most significant seems to be the observation on the position of the grid-shell's joints: apart from the boundary curve Υ_0 in all the simulations they are located only at points of the load application, in other words

$$\operatorname{spt} \mathscr{S}_h \subset \big((\operatorname{spt} \underline{F}_h) \cup \Upsilon_0\big)^2.$$

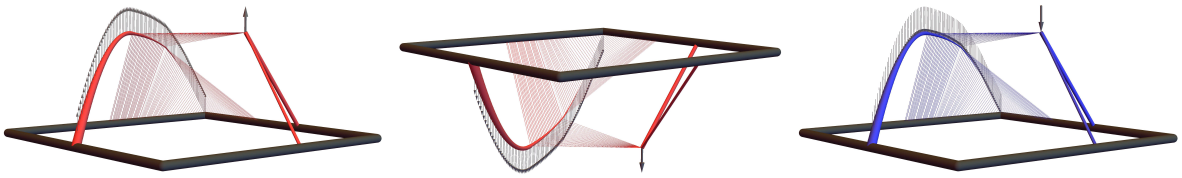
6.6. Final remarks and open problems

In the final section of this thesis let us give some additional insight into the form-finding and the Prager problems. We shall dwell on how the foregoing results can be generalized/altered and how they cannot.

6.6.1. From forms in pure tension to forms in pure compression

Let us first return to the interpretational issues pointed out in Section 6.1: in all of the examples considered so far the load f (and so the transmitted 3D load $\underline{F}_z$) were pointing upwards, i.e. $f \geq 0$, meanwhile in practise we encounter gravitational loads, which corresponds to $f \leq 0$. If $f \geq 0$ then the optimal form is to be erected above the base plane

$\mathbb{R}^d$ (see Corollary 6.19), which is visualized in Fig. 6.6(a) (it is a different view on the grid-shell approximation already displayed in Fig. 6.4). If we change the sign of the load the optimal form flips to hang below $\mathbb{R}^d$, see Fig. 6.6(b). The form thus obtained may be called *an optimal cable system*.



(a) (b) (c)

Figure 6.6: (a) Optimal grid-shell in tension (red colour) under an abstract upward load; (b) Optimal grid-shell in tension (red colour) – a cable system – under a downward gravity load; (c) Optimal grid-shell in compression (blue colour) under a downward gravity load.

All the considerations above rely on the assumption that the designed structure is in pure tension: in plastic problems (PFF) or (PPP) we explicitly demand that $\underline{\varsigma} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}_+^{\underline{d}\times\underline{d}})$ whilst in the elastic variants (EFF) or (EPP) this requirement is encoded in the energy function j_+. It is natural to address the question as follows: how the form-finding/Prager problems change if compression is enforced instead, i.e. $\underline{\tilde{\varsigma}} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}_-^{\underline{d}\times\underline{d}})$ and j_- is employed. The Prager problems remain to be well posed, all the arguments (or their mirrored counterparts) from Section 6.2.1 hold true. A similar change would be impossible for the optimal membrane problem, e.g. the convexity stated in Proposition 5.4 would be lost for j_-. Notwithstanding this the constructions of optimal forms or grid-shells from Sections 6.3, 6.4, 6.5 are still feasible provided that a change of sign is made in strategic places.

For solutions (u, w) and (ϑ, ς) of problems $(\overline{\mathscr{P}})$ and, respectively, $(\mathscr{P}^*)$ (in their original form, in particular $\varsigma \succeq 0$) we will now run through the construction of optimal form for the compression settings of form-finding/Prager problems. Let us note that in the plastic variants of the design problems the minimized functional will read $\int_{\overline{\Omega}} \underline{\rho}_-(\underline{\tilde{\varsigma}}) = -\int_{\overline{\Omega}} \operatorname{Tr} \underline{\tilde{\varsigma}}$. The optimal elevation function will be now flipped, i.e.

$$\tilde{z} := -\frac{1}{\sqrt{2}}\, u.$$

Then, for $\underline{\tilde{z}} := (\mathrm{id}, \tilde{z})$ we define the $\underline{d}$-dimensional objects

$$\underline{\tilde{\mu}} := \sqrt{2}\, \underline{\tilde{z}} \,\#\, \mu, \qquad \underline{\tilde{\sigma}} := \begin{bmatrix} -\sigma & \frac{1}{\sqrt{2}}\, q \\ \frac{1}{\sqrt{2}}\, q^\top & -\frac{1}{2}\, \langle \sigma^{-1} q, q \rangle \end{bmatrix} \circ \underline{\tilde{z}}^{-1}.$$

where μ, σ, q are defined exactly as in Section 6.3.1. Since in contrast to (6.32) we now have $\frac{1}{\sqrt{2}}\, q = -\sigma \nabla_\mu \tilde{z}$, we can see that the minus signs in definition of $\tilde{\sigma}$ are chosen so that $\tilde{\sigma} = -\big(\nabla_\mu \tilde{z}\big)^\top \sigma\, \nabla_\mu \tilde{z}$ and thus negative semi-definiteness of $\tilde{\sigma}$ follows. With $\underline{\tilde{\varsigma}} := \underline{\tilde{\sigma}}\, \underline{\tilde{\mu}}$ the pair $(\underline{\tilde{\varsigma}}, \underline{F}_{\tilde{z}})$ solves the Prager problem (PPP) in its compression variant. The Hooke function $\underline{\tilde{\mathscr{H}}}$ must $\underline{\tilde{\mu}}$-a.e. satisfy the condition similar to (6.30), although now $\underline{\sigma}(\underline{x}) = -\sum_{i=1}^m \alpha_i\, \underline{\tau}_i \otimes \underline{\tau}_i$. Then, for $\underline{\tilde{\lambda}} = \frac{C_0}{\sqrt{2}\, \mathcal{Z}}\, \underline{\tilde{\mathscr{H}}}\, \underline{\tilde{\mu}}$ the pair $(\underline{\tilde{\lambda}}, \underline{F}_{\tilde{z}})$ solves the compression variant of the elastic setting of Prager problem (EPP).

The modification towards the compressive framework is simpler when there exists a string system $(\pi, \mathscr{S})$ that is optimal for $(\widetilde{\mathscr{P}^*})$. With $\tilde{\underline{z}}$ defined as above it is enough to define the $\underline{d}$-dimensional non-positive pairing measure

$$\tilde{\underline{\mathscr{S}}} := -(\tilde{\underline{z}}, \tilde{\underline{z}}) \# \left(\sqrt{2} \left(\chi^*_{\mathrm{K}}(\alpha, 1) \right)^{\frac{1}{2}} \mathscr{S} \right)$$

and the rest of the construction that was put forth in Section 6.4 stays practically the same: the pairs $(\underline{\varsigma}_{\tilde{\underline{\mathscr{S}}}}, \underline{F}_{\tilde{z}})$ and $(\lambda_{|\tilde{\underline{\mathscr{S}}}|}, \underline{F}_{\tilde{z}})$ solve the compression variants of Prager problems (PPP) and (EPP) respectively.

The discrete version of construction of suboptimal grid-shells from Section 6.5 may be changed accordingly. The numerical prediction of an optimal form in compression may thus be given, see Fig. 6.6(c) where the blue colour is used to reflect the compressive stresses (see the colouring scheme for trusses in Fig. 4.9). We can see that the grid-shell "flips back" to the geometry configuration as in Fig. 6.6(a) for the upward load.

Finally, one may wonder what happens if in the form-finding/Prager problems we do not enforce the stress sign at all, i.e. we allow any $\underline{\varsigma} \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{\underline{d} \times \underline{d}})$ and choose the energy function j inducing the linear Hooke law. The corresponding Prager problems are well behaved just as well – solutions do exist. The connection to the optimal membrane problem, however, falls apart hence we lose the methods herein developed for showing that there are solutions of Prager problem that are actual forms. This tracks well with the papers [72, 164] where for the plane Prager problem ($d = 1$ and $\underline{d} = 2$) the authors have shown, analytically and numerically, that when both tension and compression is allowed the optimal structure is not a form, i.e. apart from an arch-like curve it also concentrates on two 2D regions.

6.6.2. A link to the optimal arch-grid problem

As was already mentioned in the introduction to this thesis, see Section 1.2.2, in developing the approach towards the optimal form-finding the present author drew inspiration from the so called *optimal arch-grid problem* that was originated at the end of '70s in [143] and recently revisited in [70], [78].

For a convex bounded domain $\Omega \subset \mathbb{R}^2$ and a fixed Cartesian basis $\mathrm{e}_1, \mathrm{e}_2$ an arch-grid can be defined as a form in compression that decomposes to two family of arches – curves lying in vertical planes whose intersections with $\mathbb{R}^2 \times \{0\}$ are straight lines parallel to e_1 (the first family) or to e_2 (the second family). Similarly as in case of the form-finding problem posed in Section 6.1 the arch-grid ought to carry the load f that is vertically transmissible.

In [70] the 3D problem of optimal arch-grids was explicitly rewritten as a pair of mutually dual convex problems posed in the planar domain $\overline{\Omega} \subset \mathbb{R}^2$. In the dual problem we seek a field $\vartheta \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^d)$ satisfying $-\operatorname{div} \vartheta = f$ in Ω to minimize a convex functional of a very specific form. In the primal problem we maximize the integral $\int_{\overline{\Omega}} u\, f$ with respect to the displacement function $u \in \mathcal{D}(\Omega; \mathbb{R})$ which must satisfy the so called *mean squared slope* condition formulated in terms of 1D integrals taken along each line parallel to e_1 and e_2, see (1.7) in Section 1.1.4. If u solves the primal problem, then the optimal arch-grid is

a form that concentrates on the surface $\overline{S}_z$ where $z = -\frac{1}{\sqrt{2}}\,u$. The latter formula has been certainly borrowed in this chapter. The reader is referred for details to the introduction in Chapter 1 herein and to the aforementioned literature.

The ideas for form-finding posed in the present chapter and for the whole methodology herein developed can be traced back to one goal: the one of generalizing the arch-gird problem by freeing the assumption that the form is composed of arches going from point on Υ_0 to another point on this curve. This was accomplished by discovering the link to the optimal membrane problem (and then by detailed analysis of this problem in cooperation with Guy Bouchitté). Mathematically this consisted in reformulation and generalization of the primal and dual problems posed for arch-grids by adding extra variables: the vector displacement function $w : \overline{\Omega} \to \mathbb{R}^2$ and, respectively, the pre-stress $\varsigma \in \mathcal{M}(\overline{\Omega}; \mathcal{S}^{2\times 2}_+)$. Optimizing with respect to such pairs: (u, w) and (ϑ, ς) is the essence of problems $(\mathscr{P}), (\mathscr{P}^*)$. The purpose of this section is to illustrate that such framework is very flexible as we can easily recover the arch-gird problem by imposing some extra constraints.

With a fixed Cartesian basis $\mathrm{e}_1, \mathrm{e}_2$, that horizontally determines the orientations of the arches, let us define a closed convex set being a modification of C defined in Section 5.2.3:

$$\mathrm{C}_{\boxplus} := \Big\{(\theta, \xi) \in \mathbb{R}^2 \times \mathcal{S}^{2\times 2} \,:\, \big\langle \tfrac{1}{2}\,\theta \otimes \theta + \xi, \mathrm{e}_i \otimes \mathrm{e}_i \big\rangle \le 1, \quad i = 1, 2\Big\}.$$

It is easy to check that $\mathrm{C} \subset \mathrm{C}_{\boxplus}$. If by $\varrho_{\boxplus}$ we denote the support function of $\mathrm{C}_{\boxplus}$ (recall that ϱ is the one for C) it can be checked that

$$\varrho^0_{\boxplus}(q, \sigma) = \varrho^0(q, \sigma) + \chi_{\mathrm{span}(\{\mathrm{e}_1\otimes\mathrm{e}_1, \mathrm{e}_2\otimes\mathrm{e}_2\})}(\sigma). \tag{6.58}$$

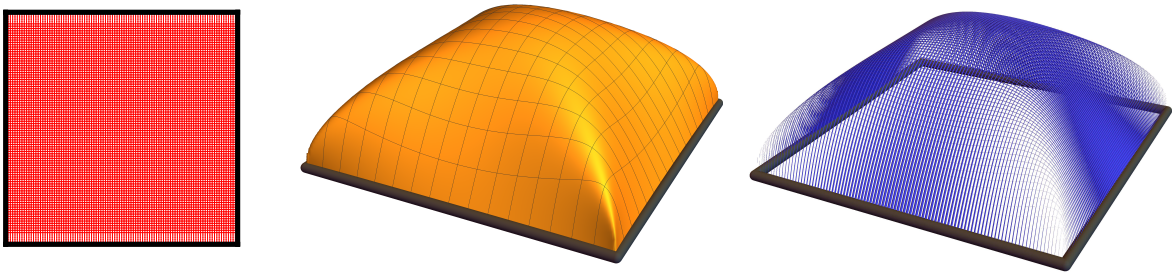
In particular $\varrho_{\boxplus}(q, \sigma) < \infty$ implies that the eigenvectors of σ are $\mathrm{e}_1, \mathrm{e}_2$ exactly (unless $\sigma = s\,\mathrm{I}$ for $s \ge 0$).

Just as the pair of problems $(\mathscr{P})$, $(\mathscr{P}^*)$ revolved around the function ϱ we now put forward the corresponding problems for $\varrho_{\boxplus}$:

$$\sup_{(u,w)\in\mathcal{D}(\Omega;\mathbb{R}\times\mathbb{R}^2)} \left\{\int u\,df \;:\; \big\langle \tfrac{1}{2}\,\nabla u \otimes \nabla u + e(w), \mathrm{e}_i \otimes \mathrm{e}_i\big\rangle \le 1 \;\text{ in } \overline{\Omega}, \;\; i = 1, 2\right\} \tag{$\mathscr{P}_{\boxplus}$}$$

$$\inf\left\{\int \varrho^0_{\boxplus}(\vartheta, \varsigma) \;:\; (\vartheta, \varsigma) \in \mathcal{M}(\overline{\Omega}; \mathbb{R}^2 \times \mathcal{S}^{2\times 2}_+),\; -\mathrm{div}\,\vartheta = f, \;\; \mathrm{Div}\,\varsigma = 0 \;\text{ in } \Omega\right\} \tag{$\mathscr{P}^*_{\boxplus}$}$$

One can show that $\mathcal{Z}_{\boxplus} := \sup \mathscr{P}_{\boxplus} = \inf \mathscr{P}^*_{\boxplus} < \infty$ and that the dual problem admits a solution ς. Considering (6.58), it must be of the form $\varsigma = s_1\,\mathrm{e}_1 \otimes \mathrm{e}_1 + s_2\,\mathrm{e}_2 \otimes \mathrm{e}_2$ for some $s_1, s_2 \in \mathcal{M}_+(\overline{\Omega})$. The in-plane equilibrium equation $\mathrm{Div}\,\varsigma = 0$ in Ω then implies a very specific form of s_1 and s_2: they decompose to segments parallel to e_1 and, respectively, e_2 (with constant density on each segment) that connect the opposing points on $\partial\Omega$. In other words, ς represents a membrane sewn out of fibres parallel to $\mathrm{e}_1, \mathrm{e}_2$ as it is demonstrated in the numerical approximation in Fig. 6.7(a) for a square domain Ω and a uniform load $f = -\mathcal{L}^2 \llcorner \Omega$. The numerical simulation itself is a straightforward modification of the method developed in Section 5.5: for the regular grid X^h as in (5.115) it is enough to build a ground structure consisting of strings parallel to $\mathrm{e}_1, \mathrm{e}_2$ only.



(a) (b) (c)

Figure 6.7: Numerical simulation of an optimal arch-grid for a square domain Ω and a uniform load $f = -\mathcal{L}^2 \llcorner \Omega$: (a) string system approximating a pre-stress ς solving $(\mathscr{P}^*_{\boxplus})$; (b) the surface $\overline{S}_{\tilde{z}_h}$; (c) the grid-shell in compression that is to approximate an optimal arch-grid.

If now one repeats the construction from Section 6.3, but this time for (u, w) and (ϑ, ς) solving $(\mathscr{P}_{\boxplus})$ and $(\mathscr{P}^*_{\boxplus})$, then we end up with a form $\underline{\varsigma}$ or $\underline{\lambda}$ with direction of material fibres that (in the top view) are restricted to align with $\mathrm{e}_1, \mathrm{e}_2$, see the numerical approximation by grid-shell in Fig. 6.7(c). This way the arch-grid-like structure is recovered, while the numerical simulation from Fig. 6.7 matches perfectly those obtained in works on optimal arch-grids, see [70], [78].

In summary: by introducing the constraints on the direction of material fibres in the optimal membrane problem we obtain 2D convex problems $(\mathscr{P}_{\boxplus})$, $(\mathscr{P}^*_{\boxplus})$ based on which we can construct the optimal arch-gird forms. Of course, the line of reasoning proposed above is far from rigorous but there is a way to connect the problem $(\mathscr{P}_{\boxplus})$ to the primal problem for the design of arch-grids given in the introduction, see (1.7). The vector function $w = w_1\,\mathrm{e}_1 + w_2\,\mathrm{e}_2$ in $(\mathscr{P}_{\boxplus})$ can be explicitly eliminated (in contrast to $(\mathscr{P})$) and the primal problem with only one variable u known for the arch-grid design is thus recast. For more details on this passage the reader is referred to the work [28, Section 8.3] by the present author.

6.6.3. Generalizations of Prager problems: when the solutions are forms?

The main statement of this chapter reads: *the $\underline{d}$-dimensional Prager problem admits a solution that is a form and such solution may be constructed explicitly from a solution of the optimal membrane problem.* Its validity, however, was rigorously proved for a very specific design setting and under some regularity assumptions. In this closing section we shall discuss the possibility of modifying the Prager problem so that form solutions remain to be optimal. For instance, based on Section 6.6.1 we already know that it is not an issue to change from the tension-only design problem to the compression-only one, while when both tension and compression is permitted then solutions of Prager problems may be other than forms. The forthcoming deliberations will lack definite answers to most of the questions. Those will be left as open problems, although we shall attempt to provide a hypothesis for some of them.

6.6.3.1. The case when solutions of $(\overline{\mathscr{P}})$ are non-Lipschitz

The construction of pairs $(\underline{\varsigma}, \underline{F}_z)$ or $(\underline{\lambda}, \underline{F}_z)$ which are forms, i.e. which satisfy $\operatorname{spt}\underline{\varsigma}$, $\operatorname{spt}\underline{\lambda} \subset \overline{S}_z$, was carried out in a fairly general setting: for any bounded star-shaped domain $\Omega \subset \mathbb{R}^d$, any load $f \in \mathcal{M}(\overline{\Omega}; \mathbb{R})$ and any height parameter $h_0 \geq \operatorname{diam}(\Omega)$. This was as far as we were able to go without further assumptions: despite having the designs $(\underline{\varsigma}, \underline{F}_z)$ and $(\underline{\lambda}, \underline{F}_z)$ at our disposal, to prove their optimality we had to assume Lipschitz continuity of the solution (u, w) of $(\overline{\mathscr{P}})$ upon which those forms are built. This leads to the natural question:

Problem 6.40. *Let us take the pairs $(\underline{\varsigma}, \underline{F}_z)$, $(\underline{\lambda}, \underline{F}_z)$ constructed as in Section 6.3.1 for arbitrary solution $(u, w) \in \overline{\mathcal{X}}_1$ of problem $(\overline{\mathscr{P}})$, namely (u, w) may not be Lipschitz continuous.*

Do the pairs $(\underline{\varsigma}, \underline{F}_z)$ and $(\underline{\lambda}, \underline{F}_z)$ solve problems (PPP) *and* (EPP)*, respectively?*

The Lipschitz continuity of the pair (u, w) was imposed so that the optimality conditions from Theorem 5.41 could be exploited, more precisely the condition $q = \sigma \nabla_\mu u$. In other words, the theory of μ-tangential calculus recalled in Section 2.5 was employed to make sure that the equilibrium equation $-\operatorname{Div}\underline{\varsigma} = \underline{F}_z$ holds in $\mathbb{R}^{\underline{d}} \setminus \underline{\Upsilon}_0$. Essentially, there are two ways to go about the problem above: either we attempt to generalize the μ-tangential calculus to any pair $(u, w) \in \overline{\mathcal{X}}_1$ or we let go off this concept and find an independent way for proving the equilibrium. The first approach has been already discussed in detail in Section 5.7.2. As far as the second strategy is concerned, the present author has not made any significant inquiry yet.

6.6.3.2. Are there solutions which are not forms?

So far this chapter tried to answer the question: do there exist optimal solutions of Prager problems (PPP), (EPP) that are forms and, if so, then how can they be constructed? A quite different problem is to settle the following: do there exist solutions $(\underline{\varsigma}, \underline{F})$, $(\underline{\lambda}, \underline{F})$ of Prager problems (PPP), (EPP) which are not forms, i.e. there is no functions $z \in C_0(\Omega)$ such that $\operatorname{spt}\varsigma, \operatorname{spt}\lambda \subset \overline{S}_z$? So far, for the Prager problems formulated exactly as in Section 6.2.1 (cf. the alterations in the forthcoming subsections), in the literature there are no examples of such non-form solutions. Consequently we ask:

Problem 6.41. *Let $(\underline{\varsigma}, \underline{F})$ (resp. $(\underline{\lambda}, \underline{F})$) be <u>any</u> solutions to the Prager problem* (PPP) *(resp.* (EPP)*).*

Does there exist a function $z \in C_0(\Omega)$ such that $\operatorname{spt}\underline{\varsigma} \subset \overline{S}_z$ (resp. $\operatorname{spt}\underline{\lambda} \subset \overline{S}_z$)?

We can go a step further and wonder whether the connection between $\underline{d}$-dimensional Prager problem and the d-dimensional optimal membrane problem is one-to-one, i.e.

Problem 6.42. *Let us take solution $(\underline{\varsigma}, \underline{F})$ or $(\underline{\lambda}, \underline{F})$ of the Prager problem* (PPP) *or* (EPP)*, respectively.*

Can the solutions be constructed in accordance with Section 6.3.1 for some solutions (u, w) and (ϑ, ς) of problems $(\overline{\mathscr{P}})$ and $(\mathscr{P}^*)$, respectively?

Since this chapter was mainly focused on the construction of the optimal form, currently the author does not know how to approach either of the two foregoing problems.

6.6.3.3. Elastic setting of Prager problem for different anisotropy settings

From the start of this chapter the elastic settings of the form-finding problem (EFF) and of the Prager problem (EPP) have been posed in the particular anisotropy setting: the one of the Canvas-like Material Design (CanMD). At the first glance the reason for this choice seems to be to achieve the equivalence to the plastic variants of those problems: (PFF) and (PPP), which is due to equality $\underline{\rho}_+(\underline{\sigma}) = \operatorname{Tr}\underline{\sigma}$ (see also Section 4.7.2 where equivalence between Michell problem and the FibMD setting of (FMD) problem was showed). On top of that, the CanMD setting was the main setting in Chapter 5 on optimal membranes thus perhaps it is not so surprising here.

The reason for the choice of CanMD setting turns out to be more profound than that. One of the keys to connecting the Prager problems to the optimal membrane problem lies in the equality $\underline{\rho}_+(\xi + \sqrt{2}\,\theta \overset{s}{\otimes} \mathrm{e}_{\underline{d}}) = \varrho(\xi, \theta)$ stated in Lemma 6.13. If the anisotropy setting of the Prager problem (EPP) changes then $\underline{\rho}_+$ changes. Then of course one can affect ϱ by changing the anisotropy setting of the Optimal Elastic Membrane problem (OEM): the natural guess would be to set the lower dimensional version of the setting chosen for (EPP) so that ρ_+ that defines ϱ has a similar structure to $\underline{\rho}_+$. However, the mysterious equality in Lemma 6.13 appears to be specific to $\underline{\rho}_+(\underline{\xi}) = \sup_{|\underline{\tau}|\leq 1} \langle \underline{\xi}, \underline{\tau} \otimes \underline{\tau} \rangle$ and its d-dimensional counterpart ρ_+. The uniqueness of this type of function also resurfaces later, at the stage of construction of a form when $\underline{\rho}^0_+(\underline{\sigma})$ is computed for the block matrix $\underline{\sigma}$ in (6.25): we find that it is equal to $\varrho^0(q, \sigma) = \rho^0_+(\sigma) + \frac{1}{2}\langle \sigma^{-1} q, q \rangle$ and all thanks to the fact that $\underline{\rho}^0_+ = \operatorname{Tr}$ and $\rho^0_+ = \operatorname{Tr}$ on the set of semi-positive matrices.

For the time being it therefore seems that methodology of this chapter cannot be generalized to the Prager or the form-finding problems in different anisotropy settings, that is for different sets $\mathcal{K}$ than $\mathcal{K}_{fib}$. Changing $\mathcal{K}$ may be important from the mechanical point of view. For instance, if the pure-compression variant (energy function j_-) is considered the choice $\mathcal{K}_{fib}$ is not as natural as for j_+ when it corresponds to design of canvas-like material. The form in compression that resists the gravitational loads should be a *vaulted structure* made of e.g. masonry (or glass in the more contemporary case). This encourages the use of $\mathcal{K} = \mathcal{K}_{iso}$, which according to Example 3.27 and Section 4.9 is called the isotropic Masonry Material Design setting. The question is as follows: do the Prager structures made of masonry material take shape of forms? More generally:

Problem 6.43. *Consider the elastic variant of the Prager problem* (EPP) *in the general anisotropy setting, i.e. for an arbitrary triple j_+ (or j_-), $\mathcal{K}$, c as in Chapter 3.*

Is there a solution $(\underline{\lambda}, \underline{F})$ that constitutes a form, i.e. does there exist a function $z \in C_0(\Omega)$ such that $\operatorname{spt}\underline{\lambda} \subset \overline{S}_z$*?*

Considering the foregoing comments this thesis completely lacks the tools to settle this issue. In the literature there are works (see e.g. [88, 56]) that present numerical results

that speak in favour of the form-like solutions, although they mainly consider the 2D case, i.e. $d = 1$, $\underline{d} = 2$.

6.6.3.4. Prager problem for an elevated supporting curve

In both the form-finding problems and the Prager problems a strong design assumption was made: that the optimal structure may rest on the horizontal curve $\underline{\Upsilon}_0 = \partial\Omega \times \{0\}$. In Corollary 6.21 we have established that (under the Lipschitz regularity assumption) the optimal forms $(\underline{\varsigma}, \underline{F}_z)$, $(\underline{\lambda}, \underline{F}_z)$ constructed for such Dirichlet region $\underline{\Upsilon}_0$ remain optimal for the Prager problem if the Dirichlet region is extended to the whole cylinder $\partial\Omega \times [-h_0, h_0]$, no matter how big h_0 is. This is remarkable, but at the same time believable considering that the load $\underline{F}$ is allowed to vertically migrate.

In terms of the form-finding problem, however it is more practical to look into a different modification of the design setting: the structure may still rest on a curve $\widetilde{\underline{\Upsilon}}_0$, but now it is defined through elevation, i.e. $\widetilde{\underline{\Upsilon}}_0 = \underline{z}_0(\partial\Omega)$ where $\underline{z}_0 = (\mathrm{id}, z_0)$ for $z_0 \in C(\partial\Omega)$. It must be stressed that z_0 is a datum, not a design variable (the latter scenario would simply give $z_0 = 0$, cf. the comment below Corollary 6.21).

Once $h_0 > \|z_0\|_{\infty,\partial\Omega}$ the Prager problems thus altered, henceforward denoted by $(\mathrm{P}\widetilde{\mathrm{P}}\mathrm{P}), (\mathrm{E}\widetilde{\mathrm{P}}\mathrm{P})$, remain to be well posed – their solutions exist. *A priori* the same cannot be told about the modified form-finding problems $(\mathrm{P}\widetilde{\mathrm{F}}\mathrm{F}), (\mathrm{E}\widetilde{\mathrm{F}}\mathrm{F})$ where now we must look for an elevation function $z \in C(\overline{\Omega})$ such that $z = z_0$ on $\partial\Omega$. Let us recall that (even for $z_0 = 0$) the form-finding problems are non-convex and they have not been directly analysed here at all. The existence showed in Corollary 6.20 and Theorems 6.31, 6.32 is by construction and it is fair to deem it a pleasant surprise. It begs the question whether this astonishing feature holds true for an arbitrary supporting curve $\widetilde{\underline{\Upsilon}}_0$, namely:

Problem 6.44. *For a function $z_0 \in C(\overline{\Omega})$ and a height parameter $h_0 > \|z_0\|_{\infty,\Omega}$ let us consider modified Prager problems $(\mathrm{P}\widetilde{\mathrm{P}}\mathrm{P}), (\mathrm{E}\widetilde{\mathrm{P}}\mathrm{P})$ in which the Dirichlet zone reads*

$$\widetilde{\underline{\Upsilon}}_0 = \underline{z}_0(\partial\Omega)$$

where $\underline{z}_0(x) = \big(x, z_0(x)\big)$.

Do there exist: function $z \in C(\overline{\Omega})$ satisfying $z = z_0$ on $\partial\Omega$ and a solution $(\underline{\varsigma}, \underline{F})$ of problem $(\mathrm{P}\widetilde{\mathrm{P}}\mathrm{P})$ (resp. solution $(\underline{\lambda}, \underline{F})$ of problem $(\mathrm{E}\widetilde{\mathrm{P}}\mathrm{P})$) such that

$$\operatorname{spt}\underline{\varsigma} \subset \overline{S}_z \quad (\textit{resp.}\ \operatorname{spt}\underline{\lambda} \subset \overline{S}_z), \qquad \underline{F} = \underline{F}_z \quad ? \tag{6.59}$$

The first attempt to settle this issue consists in adaptation of the arguments used in this chapter. We start with the estimates in Theorem 6.9 and this part ends with a success: by means of the reasoning above Corollary 6.21 we find that

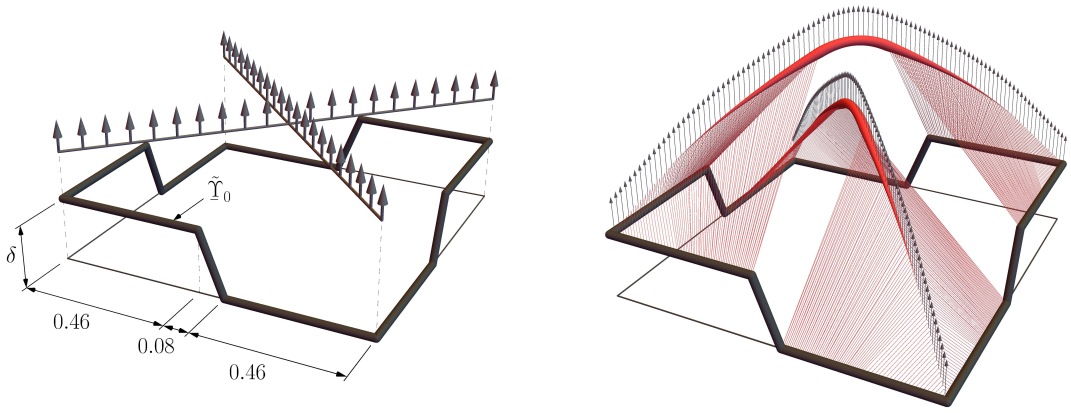
$$\inf(\mathrm{P}\widetilde{\mathrm{P}}\mathrm{P}) \geq \sqrt{2}\,\mathcal{Z}, \qquad \inf(\mathrm{E}\widetilde{\mathrm{P}}\mathrm{P}) \geq \frac{\mathcal{Z}^2}{C_0} \tag{6.60}$$

where $\mathcal{Z}$ is the value function for the optimal membrane problem (in its original form, unaffected by z_0). Indeed, the minimum in (6.37) is not greater than $\inf(\mathrm{P}\widetilde{\mathrm{P}}\mathrm{P})$ and an analogous estimate holds true for $(\mathrm{E}\widetilde{\mathrm{P}}\mathrm{P})$. Further pursuit of the construction from Section 6.3 must be quickly given up: if for a solution (u, w) of $(\overline{\mathscr{P}})$ we define $z := \frac{1}{\sqrt{2}}\,u$ we clearly

fail to realize the non-homogeneous boundary condition $z = z_0$ on $\partial\Omega$ since $u \in C_0(\Omega)$. Some attempts to modify $(\overline{\mathscr{P}})$ by enforcing boundary condition $u = \sqrt{2}\, z_0$ on $\partial\Omega$ has been undertaken by the present author but they did not allow to proceed with the construction of an optimal form.

The aforementioned failure brings us to the opposing strategy: by using inequalities (6.60) we may try to point to a counter-example. This it no easy task since we are trying to disapprove existence. Instead we shall give an example of data Ω, f, z_0 for which the solution satisfying (6.59) is "unlikely to exist":

Example 6.45. Let us consider a square domain $\Omega = (-\frac{1}{2}, \frac{1}{2})^2 \subset \mathbb{R}^2$ and the diagonal load $f = \mathcal{H}^1 \llcorner [a_1, a_3] + \mathcal{H}^1 \llcorner [a_2, a_4]$ where a_i are the corners of the square. This data was considered for the optimal membrane problem in Example 5.72, while in Fig. 5.8(a) we have the numerical approximation $\varsigma_{\mathscr{S}_h}$ of solution to $(\overline{\mathscr{P}})$. Furthermore, for h_0 sufficiently large, Fig. 6.5(c) illustrates the grid-shell approximation $(\underline{\varsigma}_{\mathscr{S}_h}, \underline{F}_h)$ of the original Prager problem (PPP), i.e. for $\Upsilon_0 = \partial\Omega \times \{0\}$. In particular $\int \operatorname{Tr} \underline{\varsigma}_{\mathscr{S}_h} = \sqrt{2}\, \mathcal{Z}_h$ and $-\operatorname{Div} \underline{\varsigma}_{\mathscr{S}_h} = \underline{F}_{z_h}$ in $\mathbb{R}^3 \setminus \Upsilon_0$ where $\underline{F}_h = \mathrm{e}_3\, z_h \# f$ while (presumably) $\mathcal{Z}_h \approx \mathcal{Z}$.



(a) (b)

Figure 6.8: Prediction of solution to the Prager problem for a non-planar supporting curve $\tilde{\Upsilon}_0$: (a) the data Ω, f and $\tilde{\Upsilon}_0$; the grid-shell approximation of solution to (P$\widetilde{\text{P}}$P).

With $\delta > 0$ chosen arbitrarily let us now pose a modified Prager problem (P$\widetilde{\text{P}}$P) for $z_0 \in C(\partial\Omega)$ that on the most part of $\partial\Omega$ takes constant values 0 or $\delta > 0$ and is affine on the other part – the resulting curve $\tilde{\Upsilon}_0$ is visualized in Fig. 6.8(a). Lacking suitable algorithms, we shall now manually construct an approximation $(\tilde{\varsigma}_h, \tilde{\underline{F}}_h)$ of solution to (P$\widetilde{\text{P}}$P).

First let us observe that the approximation $\varsigma_{\mathscr{S}_h}$ in Fig. 5.8 may be decomposed to two independent parts ς_h^1, ς_h^2, each transferring the load along the respective diagonal. In particular $\operatorname{Div} \varsigma_h^1 = 0$ and $\operatorname{Div} \varsigma_h^2 = 0$ in Ω. A similar conclusion may be drawn for the spatial grid-shell $\underline{\varsigma}_{\mathscr{S}_h}$ in Fig. 6.5(c): $\underline{\varsigma}_{\mathscr{S}_h} = \underline{\varsigma}_h^1 + \underline{\varsigma}_h^2$ while $-\operatorname{Div} \underline{\varsigma}_h^i = \mathrm{e}_3\, z_h \# f^i$ in $\mathbb{R}^3 \setminus \Upsilon_0$ for $i = 1, 2$ where $f_1 = \mathcal{H}^1 \llcorner [a_1, a_3]$ and $f_2 = \mathcal{H}^1 \llcorner [a_2, a_4]$. Next we notice that the geometry

of the elevated curve $\tilde{\Upsilon}_0$ in Fig. 6.8(a) is specifically tailored for the shape of the solution $\varsigma_{\mathscr{S}_h}$ in Fig. 5.8. We propose the construction:

$$\tilde{\varsigma}_h := \varsigma_h^1 + (\mathrm{id}, \delta) \# \varsigma_h^2$$

i.e. we modify the grid-shell in Fig. 6.5(c) by raising the suitable half by δ, see Fig. 6.8(b). We introduce a similar difference in elevation for the migrated load:

$$\tilde{\underline{F}}_h := (\mathrm{id}, z_h) \# \Big(\mathrm{e}_3 \, \mathcal{H}^1 \llcorner [a_1, a_3] \Big) \; + \; (\mathrm{id}, z_h + \delta) \# \Big(\mathrm{e}_3 \mathcal{H}^1 \llcorner [a_2, a_4] \Big).$$

Thanks to the foregoing relations in geometry we can now see that both halves ς_h^1 and $(\mathrm{id}, \delta) \# \varsigma_h^2$ now rest on the curve $\tilde{\Upsilon}_0$, i.e. $-\mathrm{Div}\, \tilde{\underline{\varsigma}}_h = \tilde{\underline{F}}_h$ in $\mathbb{R}^3 \setminus \tilde{\Upsilon}_0$ rendering the pair $(\tilde{\underline{\varsigma}}_h, \tilde{\underline{F}}_h)$ admissible for (P$\widetilde{\mathrm{P}}$P). Finally we can see that $\int \mathrm{Tr}\, \tilde{\underline{\varsigma}}_h = \int \mathrm{Tr}\, \varsigma_h^1 + \int \mathrm{Tr}\, \big((\mathrm{id}, \delta) \# \varsigma_h^2 \big) = \int \mathrm{Tr}\, \varsigma_h^1 + \int \mathrm{Tr}\, \varsigma_h^2 = \int \mathrm{Tr}\, \underline{\varsigma}_{\mathscr{S}_h} = \sqrt{2}\, \mathcal{Z}_h \approx \sqrt{2}\, \mathcal{Z}$. Considering the first inequality in (6.60) we deduce that the pair $(\tilde{\underline{\varsigma}}_h, \tilde{\underline{F}}_h)$ showed in Fig. 6.8(b) is in fact an approximation of solution to (P$\widetilde{\mathrm{P}}$P). $\triangle$

From Example 6.45 it is clear that the grid-shell $\tilde{\underline{\varsigma}}_h$ from Fig. 6.8(b) does not concentrate on a single surface $\overline{S}_z$ for any continuous function $z \in C(\overline{\Omega})$. It does not give a negative answer to Problem 6.44 yet: first of all it is merely a half-numerical prediction but, more importantly, there might be other solutions $(\tilde{\underline{\varsigma}}, \tilde{\underline{F}})$ that do meet the requirement of constituting forms. Nonetheless this seems to be very unlikely. Let us note that the jump $\delta > 0$ may be chosen arbitrarily large (provided that h_0 is big enough) while the prediction from Fig. 6.8(b) stands. It is thus difficult to believe that for δ approaching infinity the optimal structure can be built so that it is connected by a single surface.

Regardless of the answer to Problem 6.44 it is quite clear that – next to the restriction in the choice of anisotropy setting described in Section 6.6.3.3 – the necessity of taking a planar curve $\Upsilon_0 = \partial\Omega \times \{0\}$ as the Dirichlet region is yet another limitation of the methodology that was put forth in Chapter 6.

Bibliography

[1] Aage N, Andreassen E, Lazarov BS, Sigmund O (2017) Giga-voxel computational morphogenesis for structural design. Nature 550:84–86

[2] Achtziger W (1997) Topology optimization of discrete structures. In: Topology optimization in structural mechanics, Springer, pp 57–100

[3] Alberti G, Ambrosio L (1999) A geometrical approach to monotone functions in $\mathbb{R}^n$. Math Z 230:259–316

[4] Allaire G (2002) Shape optimization by the homogenization method. Springer Science & Business Media

[5] Allaire G, Kohn RV (1993) Optimal design for minimum weight and compliance in plane stress using extremal microstructures. Eur J Mech A-Solids 12:839–878

[6] Allaire G, Jouve F, Toader AM (2004) Structural optimization using sensitivity analysis and a level-set method. J Comput Phys 194:363–393

[7] Ambrosio L, Tilli P (2004) Topics on analysis in metric spaces. Oxford University Press on Demand

[8] Ambrosio L, Fusco N, Pallara D (2000) Functions of bounded variation and free discontinuity problems. Clarendon Press, Oxford

[9] Andersen ED, Roos C, Terlaky T (2003) On implementing a primal-dual interior-point method for conic quadratic optimization. Math Program 95:249–277

[10] ApS M (2019) Mosek optimization toolbox for MATLAB. User's Guide and Reference Manual, Version 4

[11] Barrett JW, Prigozhin L (2007) A mixed formulation of the Monge-Kantorovich equations. Esaim Math Model Numer Anal 41:1041–1060

[12] Beckmann M (1952) A continuous model of transportation. Econometrica 20:643–660

[13] Bellettini G, Bouchitté G, Fragalà I (1999) BV functions with respect to a measure and relaxation of metric integral functionals. J Convex Anal 6:349–366

[14] Ben-Tal A, Nemirovski A (2001) Lectures on modern convex optimization: analysis, algorithms, and engineering applications. SIAM

[15] Bendsøe MP, Haber RB (1993) The Michell layout problem as a low volume fraction limit of the perforated plate topology optimization problem: an asymptotic study. Struct Optim 6:263–267

[16] Bendsøe MP, Kikuchi N (1988) Generating optimal topologies in structural design using a homogenization method. Comput Meth Appl Mech Eng 71:197–224

[17] Bendsøe MP, Sigmund O (1999) Material interpolation schemes in topology opti-

mization. Arch Appl Mech 69:635–654

[18] Bendsøe MP, Sigmund O (2003) Topology optimization: theory, methods, and applications. Springer Science & Business Media

[19] Bendsøe MP, Sokołowski J (1995) Shape sensitivity analysis of optimal compliance functionals. Mech Struct Mach 23:35–58

[20] Bendsøe MP, Guedes JM, Haber RB, Pedersen P, Taylor JE (1994) An analytical model to predict optimal material properties in the context of optimal structural design. J Appl Mech 61:930–937

[21] Bendsøe MP, Guedes JM, Plaxton S, Taylor JE (1996) Optimization of structure and material properties for solids composed of softening material. Int J Solids Struct 33:1799–1813

[22] Bergos J (1999) Gaudí: the man and his work. Bulfinch Press

[23] Bletzinger KU, Ramm E (1999) A general finite element approach to the form finding of tensile structures by the updated reference strategy. International Journal of Space Structures 14:131–145

[24] Bletzinger KU, Wüchner R, Daoud F, Camprubí N (2005) Computational methods for form finding and optimization of shells and membranes. Comput Meth Appl Mech Eng 194(30-33):3438–3452

[25] Block P, Ochsendorf J (2007) Thrust network analysis: a new methodology for three-dimensional equilibrium. J Int Assoc Shell Spat Struct 48:167–173

[26] Bołbotowski K (2018) Theory of grillage optimization – a discrete setting. In: Rodrigues H, Herskovits J, Mota Soares C, Araújo A, Guedes J, Folgado J, Moleiro F, Madeira JFA (eds) EngOpt 2018 Proceedings of the 6th International Conference on Engineering Optimization, Springer, pp 605–618

[27] Bołbotowski K (2020) Higher order weighted Sobolev spaces on the real line for strongly degenerate weights. Application to variational problems in elasticity of beams. J Math Anal Appl 488:124038

[28] Bołbotowski K (2021) Optimal vault problem – form finding through 2D convex program. arXiv preprint arXiv:210407148

[29] Bołbotowski K, Bouchitté G (2021) Optimal design versus maximal Monge-Kantorovich metrics. To appear in: Arch Ration Mech Anal

[30] Bołbotowski K, Lewiński T (2020) Setting the Free Material Design problem through the methods of optimal mass distribution. arXiv preprint arXiv:200411084

[31] Bołbotowski K, Sokół T (2016) New method of generating Strut and Tie models using truss topology optimization. In: Advances in Mechanics: Theoretical, Computational and Interdisciplinary Issues. M. Kleiber, T. Burczyński, K. Wilde, J. Górski, K. Winkelmann, Ł. Smakosz (Eds.). CRC Press, pp 97–100

[32] Bołbotowski K, He L, Gilbert M (2018) Design of optimum grillages using layout optimization. Struct Multidiscip Optim 58:851–868

[33] Bołbotowski K, Lewiński T, Sokół T (2020) Michell structures within L-shaped domains. Computer Assisted Methods in Engineering and Science 27:185–204

[34] Bouchitté G (2020) Optimization of light structures: the vanishing mass conjecture.

arXiv preprint arXiv:200102022
[35] Bouchitté G, Buttazzo G (2001) Characterization of optimal shapes and masses through Monge-Kantorovich equation. J Eur Math Soc 3:139–168
[36] Bouchitté G, Fragalà I (2003) Second-order energies on thin structures: variational theory and non-local effects. J Funct Anal 204:228–267
[37] Bouchitté G, Fragalà I (2007) Optimality conditions for mass design problems and applications to thin plates. Arch Ration Mech Anal 184:257–284
[38] Bouchitté G, Valadier M (1988) Integral representation of convex functionals on a space of measures. J Funct Anal 80:398–420
[39] Bouchitté G, Buttazzo G, Fragalà I (1997) Mean curvature of a measure and related variational problems. Ann Scuola Norm Super Pisa-Cl Sci 25(1-2):179–196
[40] Bouchitté G, Buttazzo G, Seppecher P (1997) Energies with respect to a measure and applications to low dimensional structures. Calc Var Partial Differ Equ 5:37–54
[41] Bouchitté G, Buttazzo G, De Pascale L (2003) A p-Laplacian approximation for some mass optimization problems. J Optim Theory Appl 118:1–25
[42] Bouchitté G, Gangbo W, Seppecher P (2008) Michell trusses and lines of principal action. Math Models Meth Appl Sci 18:1571–1603
[43] Bouchitté G, Fragalà I, Seppecher P (2011) Structural optimization of thin elastic plates: the three dimensional approach. Arch Ration Mech Anal 202:829–874
[44] Bouchitté G, Fragalà I, Lucardesi I (2014) Shape derivatives for minima of integral functionals. Math Program 148:111–142
[45] Bouchitté G, Mattei O, Milton GW, Seppecher P (2019) On the forces that cable webs under tension can support and how to design cable webs to channel stresses. Proc R Soc Edinb Sect A-Math 475:2018.0781
[46] Bourdin B, Kohn RV (2008) Optimization of structural topology in the high-porosity regime. J Mech Phys Solids 56:1043–1064
[47] Braides A (2002) Gamma-convergence for beginners. Clarendon Press
[48] Brezis H (2010) Functional analysis, Sobolev spaces and partial differential equations. Springer Science & Business Media
[49] Bruggi M (2020) A constrained force density method for the funicular analysis and design of arches, domes and vaults. Int J Solids Struct 193:251–269
[50] Cannarsa P, Cardaliaguet P (2004) Representation of equilibrium solutions to the table problem for growing sandpiles. J Eur Math Soc 6:435–464
[51] Castaing C, Valadier M (1977) Convex analysis and measurable multifunctions. Springer
[52] Castaneda PP, Suquet P (1997) Nonlinear composites. Advances in Appl Mech 34:171–302
[53] Chan ASL (1960) The design of Michell optimum structures. Cranfield: The College of Aeronautics (Report 142)
[54] Chan HSY (1963) Optimum Michell frameworks for three parallel forces. Cranfield: The College of Aeronautics (Report AERO 167)
[55] Chan HSY (1967) Half-plane slip-line fields and Michell structures. Q J Mech Appl

Math 20:453–469
[56] Chiandussi G, Codegone M, Ferrero S (2009) Topology optimization with optimality criteria and transmissible loads. Comput Math Appl 57:772–788
[57] Ciarlet PG (1980) A justification of the von Kármán equations. Arch Ration Mech Anal 73:349–389
[58] Ciarlet PG (2000) Mathematical Elasticity: Volume III: Theory of Shells. North-Holland
[59] Ciarlet PG (2002) The finite element method for elliptic problems. SIAM
[60] Ciarlet PG, Rabier P (1980) Les équations de von Kármán. Springer
[61] Courant R, Hilbert D (1989) Methods of mathematical physics. Wiley-Interscience Publication
[62] Czarnecki S (2015) Isotropic material design. Computational Methods in Science and Technology 21:49–64
[63] Czarnecki S, Lewiński T (2009) On free material design problem for two loading conditions. 8th World Congresses of Structural and Multidisciplinary Optimization, Lisboa, Portugal, 1-5 June 2009. 10 pp. CD ROM. ISSMO-APMTAC
[64] Czarnecki S, Lewiński T (2012) A stress-based formulation of the free material design problem with the trace constraint and single loading condition. Bull Pol Acad Sci-Tech Sci 60:191–204
[65] Czarnecki S, Lewiński T (2014) A stress-based formulation of the free material design problem with the trace constraint and multiple load conditions. Struct Multidiscip Optim 49:707–731
[66] Czarnecki S, Lewiński T (2017) On material design by the optimal choice of Young's modulus distribution. Int J Solids Struct 110:315–331
[67] Czarnecki S, Lewiński T (2017) Pareto optimal design of non-homogeneous isotropic material properties for the multiple loading conditions. Phys Status Solidi B 254:1600821
[68] Czarnecki S, Wawruch P (2015) The emergence of auxetic material as a result of optimal isotropic design. Phys Status Solidi B 252:1620–1630
[69] Czubacki R, Lewiński T (2015) Topology optimization of spatial continuum structures made of nonhomogeneous material of cubic symmetry. J Mech Mater Struct 10:519–535
[70] Czubacki R, Lewiński T (2020) Optimal archgrids: a variational setting. Struct Multidiscip Optim 62:1371–1393
[71] Dacorogna B (2007) Direct Methods in the Calculus of Variations. Springer Science & Business Media
[72] Darwich W, Gilbert M, Tyas A (2010) Optimum structure to carry a uniform load between pinned supports. Struct Multidiscip Optim 42:33–42
[73] Day AS (1965) An introduction to dynamic relaxation. The Engineer 219:218–221
[74] Demengel F (1985) Déplacements à déformations bornées et champs de contrainte mesures. Ann Scuola Norm Super Pisa-Cl Sci 12:243–318
[75] Dorn WS, Gomory RE, Greenberg HJ (1964) Automatic design of optimal structures.

J de Mècanique 3:25–52

[76] Dweik S, Santambrogio F (2019) L^p bounds for boundary-to-boundary transport densities, and $W^{1,p}$ bounds for the BV least gradient problem in 2D. Calc Var Partial Differ Equ 58:31

[77] Dzierżanowski G (2012) Stress energy minimization as a tool in the material layout design of shallow shells. Int J Solids Struct 49:1343–1354

[78] Dzierżanowski G, Czubacki R (2021) Optimal archgrids spanning rectangular domains. Comput Struct 242:106,371

[79] Dzierżanowski GM, Wójcik-Grząba I (2020) Optimal form-finding of cable systems. Arch Civ Eng pp 305–321

[80] Ekeland I, Temam R (1999) Convex analysis and variational problems. SIAM, Oxford

[81] Evans LC, Gangbo W (1999) Differential equations methods for the Monge-Kantorovich mass transfer problem. 653, American Mathematical Soc.

[82] Evans LC, Gariepy RF (1992) Measure theory and fine properties of functions. CRC Press, Boca Raton

[83] Fabian M, Habala P, Hájek P, Montesinos V, Zizler V (2011) Banach space theory: the basis for linear and nonlinear analysis. Springer Science & Business Media

[84] Fairclough HE, Gilbert M, Pichugin AV, Tyas A, Firth I (2018) Theoretically optimal forms for very long-span bridges under gravity loading. Proc R Soc A-Math Phys Eng Sci 474:2017.0726

[85] Fox DD, Raoult A, Simo JC (1993) A justification of nonlinear properly invariant plate theories. Arch Ration Mech Anal 124:157–199

[86] Fragala I, Mantegazza C (1999) On some notions of tangent space to a measure. Proc R Soc Edinb Sect A-Math 129:331–342

[87] Fraternali F (2010) A thrust network approach to the equilibrium problem of unreinforced masonry vaults via polyhedral stress functions. Mech Res Commun 37:198–204

[88] Fuchs MB, Moses E (2000) Optimal structural topologies with transmissible loads. Struct Multidiscip Optim 19:263–273

[89] Gao DY (2018) On topology optimization and canonical duality method. Comput Meth Appl Mech Eng 341:249–277

[90] Giaquinta M, Giusti E (1985) Researches on the equilibrium of masonry structures. Arch Ration Mech Anal 88:359–392

[91] Gilbert M, Tyas A (2003) Layout optimization of large-scale pin-jointed frames. Eng Comput 20:1044–1064

[92] Gilbert M, Darwich W, Tyas A, Shepherd P (2005) Application of large-scale layout optimization techniques in structural engineering practice. 6th World Congresses of Structural and Multidisciplinary Optimization. University of Bath

[93] Gilbert M, He L, Smith CC, Le CV (2014) Automatic yield-line analysis of slabs using discontinuity layout optimization. Proc R Soc A-Math Phys Eng Sci 470:2014.0071

[94] Giusti E (1984) Minimal surfaces and functions of bounded variation. Monographs in Mathematics, 80. Birkhauser Verlag, Basel

[95] Goodman J, Kohn RV, Reyna L (1986) Numerical study of a relaxed variational problem from optimal design. Comput Meth Appl Mech Eng 57:107–127

[96] Górny W (2018) Planar least gradient problem: existence, regularity and anisotropic case. Calc Var Partial Differ Equ 57:98

[97] Górny W, Rybka P, Sabra A (2017) Special cases of the planar least gradient problem. Nonlinear Anal-Theory Methods Appl 151:66–95

[98] Graczykowski C, Lewiński T (2006) Michell cantilevers constructed within trapezoidal domains – Part I: geometry of Hencky nets. Struct Multidiscip Optim 32:347–368

[99] Graczykowski C, Lewiński T (2006) Michell cantilevers constructed within trapezoidal domains – Part II: virtual displacement fields. Struct Multidiscip Optim 32:463–471

[100] Graczykowski C, Lewiński T (2010) Michell cantilevers constructed within a half strip. Tabulation of selected benchmark results. Struct Multidiscip Optim 42:869–877

[101] Green AE, Zerna W (1968) Theoretical elasticity

[102] Haslinger J, Kočvara M, Leugering G, Stingl M (2010) Multidisciplinary free material optimization. SIAM J Appl Math 70:2709–2728

[103] He L, Gilbert M (2015) Rationalization of trusses generated via layout optimization. Struct Multidiscip Optim 52:677–694

[104] He L, Gilbert M, Johnson T, Pritchard T (2019) Conceptual design of AM components using layout and geometry optimization. Comput Math Appl 78:2308–2324

[105] Hemp WS (1973) Optimum structures. Clarendon Press, Oxford

[106] Ikeya K, Shimoda M, Shi JX (2016) Multi-objective free-form optimization for shape and thickness of shell structures with composite materials. Composite Structures 135:262–275

[107] Jiang Y, Zegard T, Baker WF, Paulino GH (2018) Form-finding of grid-shells using the ground structure and potential energy methods: a comparative study and assessment. Struct Multidiscip Optim 57:1187–1211

[108] Kantorovich LV (1948) On a problem of Monge. In: CR (Doklady) Acad. Sci. URSS (NS), vol 3, pp 225–226

[109] Kelley JL (2017) General topology. Courier Dover Publications

[110] Kočvara M, Stingl M, Zowe J (2008) Free material optimization: recent progress. Optimization 57:79–100

[111] Kohn RV, Strang G (1986) Optimal design and relaxation of variational problems, I, II, III. Comm Pure Appl Math 39:113–137, 139–182, 353–377

[112] Lechicki A (1987) Some topological problems in multifunctions theory. Rendiconti del Seminario Matematico e Fisico di Milano 57:459–463

[113] Lewinski T, Telega JJ (2000) Plates, laminates, and shells: asymptotic analysis and homogenization. World Scientific, Singapore

[114] Lewiński T, Zhou M, Rozvany GIN (1994) Extended exact solutions for least-weight truss layouts-part I: cantilever with a horizontal axis of symmetry. Int J Mech Sci 36:375–398

[115] Lewiński T, Zhou M, Rozvany GIN (1994) Extended exact solutions for least-weight truss layouts-part II: Unsymmetric cantilevers. Int J Mech Sci 36:399–419

[116] Lewiński T, Rozvany GIN, Sokół T, Bołbotowski K (2013) Exact analytical solutions for some popular benchmark problems in topology optimization III: L-shaped domains revisited. Struct Multidiscip Optim 47:937–942

[117] Lewiński T, Czubacki R, Dzierżanowski G, Sokół T (2019) Optimal archgrids revisited: variational approach and numerical methods. In: 13th World Congress of Structural and Multidisciplinary Optimization, May 20-24, Beijing, China

[118] Lewiński T, Sokół T, Graczykowski C (2019) Michell Structures. Springer International Publishing, Cham

[119] Liew A (2020) Constrained Force Density Method optimisation for compression-only shell structures. Structures 28:1845–1856

[120] Lowe PG, Melchers RE (1972) On the theory of optimal constant thickness, fibrereinforced plates. I. Int J Mech Sci 14:311–324

[121] Lu H, Gilbert M, Tyas A (2019) Layout optimization of building frames subject to gravity and lateral load cases. Struct Multidiscip Optim 60:1561–1570

[122] Lu H, Tyas A, Gilbert M, Pichugin AV (2021) On transmissible load formulations in topology optimization. Struct Multidiscip Optim 64:23–37

[123] Lurie KA, Cherkaev AV (1984) G-closure of a set of anisotropically conducting media in the two-dimensional case. J Optim Theory Appl 42:283–304

[124] Maxwell JC (1864) On reciprocal figures and diagrams of forces. Phil Mag S 27:250–261

[125] Mazurek A, Baker WF, Tort C (2011) Geometrical aspects of optimum truss like structures. Struct Multidiscip Optim 43:231–242

[126] Michael E (1956) Continuous selections. I. Ann Math pp 361–382

[127] Michell AGM (1904) The limits of economy of material in frame-structures. Phil Mag S 8:589–597

[128] Milton GW (2002) The theory of composites. Cambridge University Press

[129] de Monvel AB, Chueshov I (1998) Uniqueness theorem for weak solutions of von Kármán evolution equations. J Math Anal Appl 221:419–429

[130] Nguyen TN, Hien TD, Nguyen-Thoi T, Lee J (2020) A unified adaptive approach for membrane structures: Form finding and large deflection isogeometric analysis. Comput Meth Appl Mech Eng 369:113239

[131] Novotny AA, Sokołowski J (2020) An Introduction to the Topological Derivative Method. Springer

[132] Novotny AA, Sokołowski J, Zochowski A (2019) Topological derivatives of shape functionals. Part I, II, III. J Optim Theory Appl 180,181:341–373, 683–710, 1–22

[133] Prager W, Rozvany GIN (1977) Optimal layout of grillages. J Struct Mech 5:1–18

[134] Prager W, Shield RT (1967) A general theory of optimal plastic design. J Appl Mech 34:184–186

[135] Richardson JN, Adriaenssens S, Coelho RF, Bouillard P (2013) Coupled form-finding and grid optimization approach for single layer grid shells. Eng Struct 52:230–239

[136] Ringertz UT (1993) On finding the optimal distribution of material properties. Struct Optim 5:265–267

[137] Rockafellar R (1968) Integrals which are convex functionals. Pac J Math 24:525–539
[138] Rockafellar R (1971) Integrals which are convex functionals. II. Pac J Math 39:439–469
[139] Rockafellar RT (1969) Measurable dependence of convex sets and functions on parameters. J Math Anal Appl 28:4–25
[140] Rockafellar RT (1970) Convex analysis. Princeton University Press
[141] Rozvany GIN (1972) Grillages of maximum strength and maximum stiffness. Int J Mech Sci 14:651–666
[142] Rozvany GIN (1976) Optimal design of flexural systems: beams, grillages, slabs, plates and shells. Pergamon Press, London
[143] Rozvany GIN, Prager W (1979) A new class of structural optimization problems: optimal archgrids. Comput Meth Appl Mech Eng 19:127–150
[144] Rozvany GIN, Sokół T (2012) Exact truss topology optimization: allowance for support costs and different permissible stresses in tension and compression - extensions of a classical solution by Michell. Struct Multidiscip Optim 45:367–376
[145] Rozvany GIN, Wang CM (1983) On plane Prager structures I. Int J Mech Sci 25:519–527
[146] Rozvany GIN, Wang CM, Dow M (1982) Prager-structures: archgrids and cable networks of optimal layout. Comput Meth Appl Mech Eng 31:91–113
[147] Rudin W (1991) Functional analysis. McGraw-Hill, New York
[148] Rybka P, Sabra A (2019) The planar Least Gradient problem in convex domains, the case of continuous datum. To appear in: Nonlinear Anal-Theory Methods Appl
[149] Rybka P, Sabra A (2021) The planar Least Gradient problem in convex domains: the discontinuous case. NoDea-Nonlinear Differ Equ Appl 28:15
[150] Rybka P, Zatorska-Goldstein A (2020) A stationary heat conduction problem in low dimensional sets in $\mathbb{R}^n$. Calc Var Partial Differ Equ 59:40
[151] Saint-Pierre J (1985) Point de Steiner et sections lipschitziennes. Sém Anal Convexe 15:1–42
[152] Santambrogio F (2015) Optimal transport for applied mathematicians. Birkäuser/Springer, Cham
[153] Save M, Prager W (1985) Structural optimization. Vol 1. Optimality criteria. Plenum Press, New York and London
[154] Schek HJ (1974) The force density method for form finding and computation of general networks. Comput Meth Appl Mech Eng 3:115–134
[155] Schneider R (2014) Convex bodies: the Brunn–Minkowski theory. Cambridge University Press
[156] Smirnov SK (1993) Decomposition of solenoidal vector charges into elementary solenoids, and the structure of normal one-dimensional flows. Algebra i Analiz 5:206–238
[157] Sokół T (2011) A 99 line code for discretized Michell truss optimization written in Mathematica. Struct Multidiscip Optim 43:181–190
[158] Sokół T (2015) Multi-load truss topology optimization using the adaptive ground

structure approach. In: Lodygowski T, Rakowski J, Litewka P (Eds.), Recent advances in computational mechanics, pp 9–16

[159] Sokół T, Lewiński T (2010) On the solution of the three forces problem and its application in optimal designing of a class of symmetric plane frameworks of least weight. Struct Multidiscip Optim 42:835–853

[160] Sokołowski J (1996) Displacement derivatives in shape optimization of thin shells. Research Report RR-2995, INRIA pp21 inria-00073702

[161] Sokołowski J, Zolésio JP (1992) Introduction to shape optimization. Springer

[162] Strang G, Kohn RV (1983) Hencky-Prandtl nets and constrained Michell trusses. Comput Meth Appl Mech Eng 36:207–222

[163] Stromberg LL, Beghini A, Baker WF, Paulino GH (2011) Application of layout and topology optimization using pattern gradation for the conceptual design of buildings. Struct Multidiscip Optim 43:165–180

[164] Tyas A, Pichugin AV, Gilbert M (2011) Optimum structure to carry a uniform load between pinned supports: exact analytical solution. Proc R Soc A-Math Phys Eng Sci 467:1101–1120

[165] Venturini S (1989) Pseudodistances and pseudometrics on real and complex manifolds. Ann Mat Pura Appl 154:385–402

[166] Villani C (2003) Topics in optimal transportation. American Mathematical Soc.

[167] Vouga E, Höbinger M, Wallner J, Pottmann H (2012) Design of self-supporting surfaces. ACM Trans Graph 31:1–11

[168] Weldeyesus AG, Stolpe M (2015) A primal-dual interior point method for large-scale free material optimization. Comput Optim Appl 61:409–435

[169] Weldeyesus AG, Stolpe M (2016) Free material optimization for laminated plates and shells. Struct Multidiscip Optim 53:1335–1347

[170] Werner R (2000) Free material optimization, PhD thesis. Friedrich–Alexander University of Erlangen, Nuremberg

[171] Zalinescu C (2002) Convex analysis in general vector spaces. World Scientific

[172] Zegard T, Paulino GH (2014) GRAND-ground structure based topology optimization for arbitrary 2D domains using MATLAB. Struct Multidiscip Optim 50:861–882

[173] Zowe J, Kočvara M, Bendsøe MP (1997) Free material optimization via mathematical programming. Math Program 79:445–466