

Quasi-Arithmetic Filters for Topology Optimization

Linus Hägg



Licentiate thesis, 2016

Department of Computing Science

This work is protected by the Swedish Copyright Legislation (Act 1960:729)

ISBN: 978-91-7601-409-7

ISSN: 0348-0542

UMINF 16.04

Electronic version available at <http://umu.diva-portal.org>

Printed by: Print & Media, Umeå University, 2016

Umeå, Sweden 2016

Acknowledgments

I am grateful to my scientific advisors Martin Berggren and Eddie Wadbro for introducing me to the fascinating subject of topology optimization, for sharing their knowledge in countless discussions, and for helping me improve my scientific skills. Their patience in reading and commenting on the drafts of this thesis is deeply appreciated. I look forward to continue with our work.

Without the loving support of my family this thesis would never have been finished. Especially, I am thankful to my wife Lovisa for constantly encouraging me, and covering for me at home when needed.

I would also like to thank my colleges at the Department of Computing Science, UMIT research lab, and at SP Technical Research Institute of Sweden for providing a most pleasant working environment.

Finally, I acknowledge financial support from the Swedish Research Council (grant number 621-3706), and the Swedish strategic research programme eSENCE.

Linus Hägg
Skellefteå 2016

Abstract

Topology optimization is a framework for finding the optimal layout of material within a given region of space. In material distribution topology optimization, a material indicator function determines the material state at each point within the design domain. It is well known that naive formulations of continuous material distribution topology optimization problems often lack solutions. To obtain numerical solutions, the continuous problem is approximated by a finite-dimensional problem. The finite-dimensional approximation is typically obtained by partitioning the design domain into a finite number of elements and assigning to each element a design variable that determines the material state of that element. Although the finite-dimensional problem generally is solvable, a sequence of solutions corresponding to ever finer partitions of the design domain may not converge; that is, the optimized designs may exhibit mesh-dependence. Filtering procedures are amongst the most popular methods used to handle the existence issue related to the continuous problem as well as the mesh-dependence related to the finite-dimensional approximation. Over the years, a variety of filters for topology optimization have been presented.

To harmonize the use and analysis of filters within the field of topology optimization, we introduce the class of *fW-mean filters* that is based on the weighted quasi-arithmetic mean, also known as the weighted generalized *f*-mean, over some neighborhoods. We also define the class of *generalized fW-mean filters* that contains the vast majority of filters for topology optimization. In particular, the class of *generalized fW-mean filters* includes the *fW-mean filters*, as well as the *projected fW-mean filters* that are formed by adding a projection step to the *fW-mean filters*.

If the design variables are located in a regular grid, uniform weights are used within each neighborhood, and equal sized polytope shaped neighborhoods are used, then a cascade of *generalized fW-mean filters* can be applied with a computational complexity that is linear in the number of design variables. Detailed algorithms for octagonal shaped neighborhoods in 2D and rhombicuboctahedron shaped neighborhoods in 3D are provided. The theoretically obtained computational complexity of the algorithm for octagonal shaped neighborhoods in 2D has been numerically verified. By using the same type of algorithm as for filtering, the additional computational complexity for computing derivatives needed in gradient based optimization is also linear in the number of design variables.

To exemplify the use of *generalized fW-mean filters* in topology optimization, we consider minimization of compliance (maximization of global stiffness) of linearly elastic continuum bodies. We establish the existence of solutions to a version of the continuous minimal compliance problem when a cascade of *projected continuous fW-mean filters* is included in the formulation. Bourdin's classical existence result for the linear density filter is a partial case of this general theorem for *projected continuous fW-mean filters*. Inspired by the works of Svanberg & Svård and Sigmund, we introduce the harmonic open-close filter, which is a cascade of four *fW-mean filters*. We present large-scale numerical experiments indicating that, for minimal compliance problems, the harmonic open-close filter produces almost binary designs, provides independent size control on both material and void regions, and yields mesh-independent designs.

Sammanfattning

Strukturoptimering är ett samlingsnamn för ett flertal metoder som syftar till att utforma elastiska strukturer på ett optimalt sätt med avseende på exempelvis hållfasthet och vikt. Det på förhand bestämda område som begränsar den elastiska strukturens utbredning brukar kallas för designdomänen. Bland strukturoptimeringsmetoderna är topologi-optimering den mest allmänna. I topologi-optimering bestäms förutom storlek och form även topologiska egenskaper. Till materiefördelningens topologiska egenskaper räknas hur många hål den innehåller samt hur dess olika delar hänger samman med varandra. Det är just förmågan att optimera de topologiska egenskaperna som skiljer topologi-optimering från mindre allmänna strukturoptimeringsmetoder såsom storleks- eller randformsop-timering. Materietillståndet i varje punkt i designdomänen beskrivs med hjälp av en funktion som kallas materieindikeringsfunktion. Målet med topologi-optimeringen är således att bestämma den optimala materieindikeringsfunktionen. I allmänhet saknar dock det resulterande optimeringsproblemet lösning, såvida inte ytterligare åtgärder vidtas.

I syfte att generera numeriska lösningar kan det oändligtdimensionella problemet approx-imeras med ett ändligtdimensionellt problem. Den ändligtdimensionella approximationen erhålls genom att indela designdomänen i ett ändligt antal delområden och därmed blir målet att optimera varje delområdes materietillstånd. Fördelen med det ändligtdimen-sionella problemet är att detta i allmänhet är lösbart. Emellertid saknas garanti för att en följd av lösningar, hörande till allt finare indelningar av designdomänen, konvergerar. Det som typiskt händer är att när indelningen av designdomänen förfinas uppenbar sig allt finare strukturer i den optimerade materiefördelningen. Topologi-optimeringsproblem som betar sig på detta sätt brukar kallas nätberoende. Användningen av filter hör till de mest populära metoderna för att hantera lösbarteproblematiken hos det oändligtdimensionella problemet, såväl som för att hantera nätberoende hörande till den ändligtdimensionella approximationen. Något förenklat kan man se det som att små strukturer filtreras bort. Idén att använda filter för topologi-optimering kommer ursprungligen från bildanalysen, där filter till exempel används för att reducera brus i bilder. Under årens lopp har ett stort antal filter med olika egenskaper föreslagits för topologi-optimering.

Med syftet att harmonisera användningen av filter inom topologi-optimering har vi introducerat klassen av *fW-medelvärdesfilter* som bygger på de kvasiaritmetiska medelvärdena. Mer precist beräknas för varje delområde i indelningen av designdomänen ett viktat kvasiaritmetiskt medelvärde över delområdets grannskap. Med ett delområdes grannskap menas en samling av till detta delområde närliggande delområden. De kvasiar-itmetiska medelvärdena liknar på många sätt det vanliga aritmetiska medelvärdet men med det viktiga undantaget att de kvasiaritmetiska i allmänhet är icke-linjära. För att kunna beskriva de flesta filtren som används för topologi-optimering har vi även introduc-erat klassen av *generaliserade fW-medelvärdesfilter*. Speciellt innefattas de redan nämnda *fW-medelvärdesfiltren* och även så kallade *projicerade fW-medelvärdesfilter* i denna klass. För att erhålla ytterligare filter med bättre egenskaper kan ett antal filter kombineras i en kaskad.

Om designdomänen indelas på ett reguljärt sätt (exempelvis i ett rutnät), om alla delområden inom ett givet grannskap tilldelas samma vikt och om alla grannskap är formade som en och samma polytop (flerdimensionell motsvarighet till månghörningarna) så växer antalet aritmetiska operationer (+, −, ·, /) som krävs för att applicera en kaskad av generaliserade fW -medelvärdesfilter proportionellt med antalet delområden i indelningen. Antalet aritmetiska operationer är ett mått på algoritmens komplexitet och kan användas för att uppskatta tiden det tar att applicera filtret. Den relativt sett låga komplexiteten för den föreslagna algoritmen öppnar upp för större (finare indelningar) problem samt för filterkaskader som tidigare ansetts för krävande. Vi beskriver i detalj algoritmen för specialfallen med oktagonformade grannskap i 2D och rombkuboktaederformade grannskap i 3D. Den teoretiska uppskattningen av komplexiteten för algoritmen i fallet med oktagonformade grannskap i 2D har verifierats numeriskt genom att mäta hur beräkningstiden beror på antalet delområden i indelningen av designdomänen. Om optimeringsproblemet är sådant att det är möjligt att beräkna derivator, så kan effektiva gradientbaserade optimeringsalgoritmer användas för dess lösning. Vi visar att om samma typ av algoritm används som vid filtreringen så är den extra komplexiteten, relaterad till en kaskad av generaliserade fW -medelvärdesfilter, för beräkning av derivator också proportionell mot antalet delområden i indelningen av designdomänen.

Ett klassiskt topologioptimeringsproblem är att finna den materiefördelning som under en given last uppvisar störst styvhet (minst eftergivlighet, så kallad komplians). Materialet antas vara linjärt elastiskt, vilket innebär att de uppkomna deformationerna står i proportion till den påförda lasten samt att materialet återtar sin ursprungliga form om lasten avlägsnas. Utan begränsning av materiefördelningens maximala volym maximeras styvheten då hela designdomänen fylls med material. För att göra problemet intressant införs därför en begränsning av materiefördelningens maximala volym. Vi bevisar att en kaskad av projicerade fW -medelvärdesfilter kan användas för att garantera existens av lösningar till det oändligtdimensionella problemet. Inspirerade av andras tidigare arbeten introducerar vi det harmoniska open–close filtret som är en kaskad av fyra fW -medelvärdesfilter. Vi presenterar storskaliga numeriska experiment som indikerar att det ändligtdimensionella problemet är nätoberoende när det harmoniska open–close filtret används. Vidare indikerar de numeriska resultaten att detta filter gör det möjligt att kontrollera storleken på de minsta strukturerna och de minsta hålen i materiefördelningen oberoende av varandra. Ur ett tillverkningsperspektiv är detta en mycket önskvärd egenskap.

List of Papers

This thesis is based on the following papers:

- I Eddie Wadbro and Linus Hägg, *On quasi-arithmetic mean based filters and their fast evaluation for large-scale topology optimization*. Structural and Multidisciplinary Optimization, 52(5):879–888, 2015, doi:10.1007/s00158-015-1273-5.
- II Linus Hägg and Eddie Wadbro, *Nonlinear filters in topology optimization: Existence of solutions and efficient implementation for minimal compliance problems*. Department of Computing Science, Umeå University, Technical report UMINF 16.01, 2016.

In addition to the papers included in thesis the following report was completed during the studies:

Linus Hägg, Eddie Wadbro, Daniel Noreland and Martin Berggren, *1D-model of the interaction between a stack of wood and an imposed electromagnetic wave*. Department of Computing Science, Umeå University, Technical report UMINF 16.03, 2016.

Contents

1	Introduction	1
2	Material distribution topology optimization	2
3	Optimization of linearly elastic continuum structures	4
3.1	Linear elastostatics	4
3.1.1	Variational form of linear elastostatics	6
3.2	Compliance minimization	6
4	Density filters in topology optimization	11
5	Thesis contributions	16
5.1	General filters and the class of fW -mean filters	16
5.2	fW -mean filtered compliance minimization	19
	Bibliography	23
	Paper I	
	Paper II	

1. Introduction

A central problem in engineering is to determine how to design a device with desirable performance. As the problem is formulated, it is likely that there exists not just one, but many designs that would satisfy our demands on the device. It is then natural to wonder whether one can find a design that is optimal in some sense. Depending on the situation, optimality could for instance be with respect to performance or manufacturing costs. If we have a mathematical model that allows us to accurately predict the performance of our devices, we can optimize without having to manufacture and experimentally evaluate a large number of prototypes.

The complexity of finding the optimal design is related to the number of designs that satisfy the demands. In most cases, it is not tractable to enumerate all feasible designs, evaluate their performance, and choose the best one. As a guide towards an optimal design, we can use mathematical programming techniques. The strategy of a typical optimization algorithm is to iteratively update the design in order to improve the performance. We note the resemblance of the iterative strategy of the algorithm and the methodology of a human designer.

Design optimization methods are often classified by the generality of the update used to modify designs, ranging from *sizing optimization*, via *boundary shape optimization* to *topology optimization* in order of increasing generality. In sizing optimization, we aim at determining the optimal sizes of the constitutive parts of a given structure, for instance optimizing the diameters of the bars in a given truss network. Boundary shape optimization amounts to determining not only the optimal sizes but also the optimal shapes of the constitutive parts of a given structure by displacing the boundaries. Typically, the connectedness of the optimized design resulting from boundary shape optimization is the same as that of the initial design. Topology optimization is by far the most general design optimization method, in which also the connectedness of the design is optimized. Often topology optimization is used to find a good conceptual design that is further improved by using boundary shape or sizing optimization.

Since Bendsøe & Kikuchi [2] introduced the *material distribution method* for the design of elastic continuum structures in 1988, the field of topology optimization has been subject to intense research. Today, the material distribution method has been successfully employed for topology optimization problems originating from a variety of different physical disciplines. A comprehensive account on topology optimization and its various applications can be found in the monograph by Bendsøe & Sigmund [3] or in the more recent reviews by Sigmund & Maute [15] and Deaton & Grandhi [8].

2. Material distribution topology optimization

Topology optimization addresses the problem of finding the distribution of material within a given region of space that yields the best performance. The word topology refers to that the explored designs not only differ with respect to size and shape but also with respect to connectedness. The following is an abstract topology optimization problem for optimizing the layout of a material with fixed properties.

$$\text{Find } \Omega^* \in \mathcal{A} \text{ such that } J(\Omega^*) = \inf_{\Omega \in \mathcal{A}} J(\Omega), \quad (2.1)$$

where \mathcal{A} denotes the *set of feasible designs* consisting of those subsets of the *design domain* $\Omega_D \subset \mathbb{R}^d$ that satisfy all required constraints, and $J : \mathcal{A} \rightarrow \mathbb{R}$ is the *objective function* to be minimized. Typically, the problem (2.1) includes equations describing the physical state of the design under a specified set of data. Examples of constraints defining \mathcal{A} are restrictions on minimal feature and minimal hole sizes, which are imposed to ensure manufacturability of the optimized designs. Without knowing more about the properties of \mathcal{A} and J , we cannot in general guarantee that the abstract problem (2.1) is well-posed. This means in particular that there can exist non-convergent minimizing sequences. To resolve any existence issue one can make a suitable reduction or enlargement of \mathcal{A} , and this is referred to as *restriction* and *relaxation*, respectively [4].

Different ways of representing the designs in \mathcal{A} give rise to different methods for solving topology optimization problems. The most common choice is to use a *material indicator function* $\rho : \Omega_D \rightarrow \{0, 1\}$ to indicate the absence ($\rho = 0$) or presence ($\rho = 1$) of material within the design domain [3].

To obtain a numerical solution, the design domain Ω_D is typically partitioned into n elements and the aim of the optimization is to determine the *design vector* $\boldsymbol{\rho} \in \{0, 1\}^n$ that indicates the presence or absence of material within each element. The discretization of the problem results in a nonlinear integer optimization problem that is computationally expensive to solve, especially for large-scale problems.

To enable the use of gradient based optimization algorithms, which are efficient for handling a large number of design variables, the discretized problem may be relaxed by allowing for intermediate values $\boldsymbol{\rho} \in [0, 1]^n$. For the relaxation to make sense, we introduce a bijective *material interpolation function* $m : [0, 1] \rightarrow \mathcal{M}$ that interpolates between the material properties $m_0 \in \mathcal{M}$ of the medium that surrounds the design and those of our fixed material $m_1 \in \mathcal{M}$ such that $m(0) = m_0$ and $m(1) = m_1$. For instance, in solid mechanics applications the design is typically assumed to be surrounded by void. Using the material interpolation function, the material properties of element i is given by $m(\rho_i)$ for $\rho_i \in [0, 1]$.

An artifact of the relaxation is that the optimized designs are not in general binary and hence not so easy to interpret. However, by using a *penalization technique*, almost binary designs are promoted. Unfortunately, an optimized design resulting from the relaxed and penalized problem is typically mesh dependent; that is, the solution depends strongly on the particular partition used to discretize Ω_D . The mesh dependence can sometimes be traced to lack of existence of solutions to the corresponding continuous problem. Borrvall [4]

presents a systematic review on several techniques that aims at resolving the issue of mesh-dependency.

One of the most popular strategies to achieve mesh-independent solutions is to use a *filtering procedure* to either modify the design variables or the derivatives of the objective function. The former type of filtering procedure is referred to as *density filtering* [5, 6] and the latter *sensitivity filtering* [13]. (It is common practice in the field of structural optimization to refer to the material indicator function as the density.) In the case of a density filter $\mathbf{F} : [0, 1]^n \rightarrow [0, 1]^n$, the material properties of element i is given by $m(F_i(\boldsymbol{\rho}))$ and $\mathbf{F}(\boldsymbol{\rho})$ is often termed the *physical design*. The reason to use the term physical design is that $\mathbf{F}(\boldsymbol{\rho})$ determines the physical behavior of the design. This means that $\mathbf{F}(\boldsymbol{\rho})$, rather than $\boldsymbol{\rho}$, should serve as a blueprint for manufacturing. For the same reason any constraints on the volume or mass of the design should be imposed on $\mathbf{F}(\boldsymbol{\rho})$.

3. Optimization of linearly elastic continuum structures

In this section, we introduce the archetypal topology optimization problem of minimizing the compliance of a linearly elastic continuum structure under static loading. To state the problem, we first need to examine the equations of linear elastostatics. A more detailed account on elasticity can, for instance, be found in Gurtin [10].

3.1 Linear elastostatics

A material that deforms under a load and resumes its undeformed shape when the load is removed is called elastic. If the load-induced deformations are small, many solids are elastic, and the relationship between the applied load and the resulting deformation is linear. Hence, many materials behave linearly elastic when the load-induced deformations are small. Linear elastostatics is the theory describing the equilibrium deformation and internal stress distribution of a linearly elastic solid under a given static load in the limit of small deformations. The region occupied by the undeformed body is often referred to as its reference configuration. We now proceed by a review of the equations of linear elasticity.

Let $\tilde{\Omega} \subset \mathbb{R}^d$ be the static equilibrium configuration of a linearly elastic body under a load, and denote its unloaded reference configuration by $\Omega \subset \mathbb{R}^d$, where d is the number of spatial dimensions. The equilibrium assumption implies that the forces acting on any sub body must balance, that is,

$$0 = \int_{\omega} \mathbf{b} + \int_{\partial\omega} \mathbf{t}(\mathbf{n}), \quad (3.1)$$

where $\omega \subset \tilde{\Omega}$ is the region occupied by the sub body, \mathbf{b} is the volume force density acting on the body, $\mathbf{t}(\mathbf{n})$ is the surface force density acting on the boundary $\partial\omega$ of the sub body, and \mathbf{n} is the outward unit normal vector to $\partial\omega$. Since the type of measure is evident from the domain of integration, whenever there is no risk of confusion, we omit the measure symbol in the integrals. Cauchy's theorem [10, p. 101] states that the surface force density $\mathbf{t}(\mathbf{n})$ depends linearly on \mathbf{n} , that is, $t_i(\mathbf{n}) = \sigma_{ij}n_j$, where σ_{ij} are the components of the second order stress tensor $\boldsymbol{\sigma}$. Throughout this thesis, we use the Einstein summation convention; that is, unless otherwise stated, in any indexed term, summation is implied for all indices occurring twice. Using that $\mathbf{t}(\mathbf{n}) = \boldsymbol{\sigma}\mathbf{n}$ together with the divergence theorem, force balance (3.1), and that ω is an arbitrary subdomain of $\tilde{\Omega}$, we conclude that

$$-\nabla \cdot \boldsymbol{\sigma} = \mathbf{b} \text{ in } \tilde{\Omega}. \quad (3.2)$$

Apart from force balance, the assumption of static equilibrium also requires that the torques acting on an arbitrary sub body must balance. By a similar argument that led to (3.2), the balance of torques implies that

$$\boldsymbol{\sigma}^T = \boldsymbol{\sigma} \text{ in } \tilde{\Omega}; \quad (3.3)$$

that is, the stress tensor must be symmetric. Due to the assumption of small deformations, equations (3.2), (3.3), and the applied boundary conditions can be assumed to hold on the reference configuration Ω . The displacement field \mathbf{u} is defined by

$$\tilde{\Omega} = \{\mathbf{x} + \mathbf{u}(\mathbf{x}) \mid \mathbf{x} \in \Omega\} \quad (3.4)$$

and provides a link between the deformed and undeformed configurations of the body. In linear elasticity, deformations are characterized by the second order symmetric (infinitesimal) strain tensor

$$\boldsymbol{\varepsilon}(\mathbf{u}) = \frac{1}{2}(\nabla\mathbf{u} + \nabla\mathbf{u}^T); \quad (3.5)$$

that is, the symmetric part of $\nabla\mathbf{u}$. The skew-symmetric part of the displacement gradient $(\nabla\mathbf{u} - \nabla\mathbf{u}^T)/2$ is related to rigid rotations and hence not describing deformations of the body.

The linear relationship between the applied load and the deformation is given by Hooke's generalized law

$$\boldsymbol{\sigma} = \mathbf{E}\boldsymbol{\varepsilon}(\mathbf{u}), \quad (3.6)$$

where \mathbf{E} is the fourth order elasticity tensor having in total d^4 components. However, the number of independent components can be reduced by invoking the symmetries [10, § 29]

$$E_{ijkl} = E_{jikl} = E_{ijlk} = E_{klij}. \quad (3.7)$$

The elastic energy density u associated with the deformation of the body is given by

$$u = \frac{1}{2}\boldsymbol{\sigma} : \boldsymbol{\varepsilon} = \frac{1}{2}\boldsymbol{\varepsilon} : \mathbf{E}\boldsymbol{\varepsilon} = \frac{1}{2}\varepsilon_{ij}E_{ijkl}\varepsilon_{kl}. \quad (3.8)$$

Since the energy density must be positive for all nonzero strains, the elasticity tensor must be positive definite in the following sense. There exists a constant $\mu > 0$ such that

$$\mathbf{S} : \mathbf{E}\mathbf{S} \geq \mu\mathbf{S} : \mathbf{S} \text{ for all } \mathbf{S}^T = \mathbf{S}. \quad (3.9)$$

We assume that the body is clamped at a non-empty open part of $\partial\Omega$ denoted by Γ_D and subject to a surface load \mathbf{t} at the rest of the boundary, $\Gamma_L = \partial\Omega \setminus \bar{\Gamma}_D$. With these boundary conditions, we finally arrive at the boundary value problem

$$-\nabla \cdot (\mathbf{E}\boldsymbol{\varepsilon}(\mathbf{u})) = \mathbf{b} \text{ in } \Omega, \quad (3.10a)$$

$$\mathbf{u} = \mathbf{0} \text{ on } \Gamma_D, \quad (3.10b)$$

$$(\mathbf{E}\boldsymbol{\varepsilon}(\mathbf{u}))\mathbf{n} = \mathbf{t} \text{ on } \Gamma_L, \quad (3.10c)$$

which forms the basis for topology optimization of linearly elastic continuum structures. The predominant choice for generating numerical solutions to boundary value problem (3.10) is the finite element method. The finite element method relies on a variational form of boundary value problem (3.10), which formally is obtained by multiplying equation (3.10a) by a test function, performing integration by parts over Ω , and invoking boundary conditions (3.10b) and (3.10c).

3.1.1 Variational form of linear elastostatics

As in the previous section, let $\Omega \subset \mathbb{R}^d$ denote the unloaded reference configuration of a linearly elastic continuum body. We assume that Ω is a bounded, connected, and open domain with Lipschitz boundary $\partial\Omega$. Moreover, we assume that $\mathbf{b} \in L^2(\Omega)^d$ and $\mathbf{t} \in L^2(\Gamma_L)^d$, and introduce the space of kinematically admissible displacements

$$\mathcal{U} = \{\mathbf{u} \in H^1(\Omega)^d \mid \mathbf{u}|_{\Gamma_D} \equiv \mathbf{0}\}, \quad (3.11)$$

equipped with the norm

$$\|\cdot\|_{\mathcal{U}} = \|\cdot\|_{H^1(\Omega)^d}. \quad (3.12)$$

By $L^2(A)$, we denote the space of square-integrable functions on A , where A is an open set in \mathbb{R}^d or a part of the boundary of such a set. The space of square-integrable functions on Ω with (weak) derivatives that also are square integrable is denoted by $H^1(\Omega)$.

The equilibrium displacement of the body is the solution of the following variational form of boundary value problem (3.10).

$$\text{Find } \mathbf{u} \in \mathcal{U} \text{ such that } a(\mathbf{u}, \mathbf{v}) = \ell(\mathbf{v}) \text{ for all } \mathbf{v} \in \mathcal{U}, \quad (3.13)$$

where the energy bilinear form $a : \mathcal{U}^2 \rightarrow \mathbb{R}$ and the load linear form $\ell : \mathcal{U} \rightarrow \mathbb{R}$ are given by

$$a(\mathbf{u}, \mathbf{v}) = \int_{\Omega} \mathbf{E}\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v}) = \int_{\Omega} E_{ijkl}\varepsilon_{ij}(\mathbf{u})\varepsilon_{kl}(\mathbf{v}), \quad (3.14)$$

$$\ell(\mathbf{v}) = \int_{\Omega} \mathbf{b} \cdot \mathbf{v} + \int_{\Gamma_L} \mathbf{t} \cdot \mathbf{v}. \quad (3.15)$$

It can be shown that both $a(\cdot, \cdot)$ and $\ell(\cdot)$ are continuous; that is, there exist constants C_1 and C_2 such that

$$|a(\mathbf{u}, \mathbf{v})| \leq C_1 \|\mathbf{u}\|_{\mathcal{U}} \|\mathbf{v}\|_{\mathcal{U}} \text{ for all } \mathbf{u}, \mathbf{v} \in \mathcal{U}, \quad (3.16)$$

$$|\ell(\mathbf{v})| \leq C_2 \|\mathbf{v}\|_{\mathcal{U}} \text{ for all } \mathbf{v} \in \mathcal{U}. \quad (3.17)$$

The positive definiteness (3.9) of \mathbf{E} in combination with Korn's inequality [7, Theorem 6.15-4], yields that $a(\cdot, \cdot)$ is \mathcal{U} -coercive; that is, there exists a positive constant C_3 such that

$$a(\mathbf{v}, \mathbf{v}) \geq C_3 \|\mathbf{v}\|_{\mathcal{U}}^2 \text{ for all } \mathbf{v} \in \mathcal{U}. \quad (3.18)$$

Since \mathcal{U} is a closed subspace of $H^1(\Omega)^d$ all assumptions of the Lax–Milgram lemma [7, Theorem 6.2-1] are met, and variational problem (3.13) is uniquely solvable. Solutions to (3.13) are called weak solutions to the original boundary value problem (3.10).

3.2 Compliance minimization

In this section, we present a version of the minimal compliance problem suitable for material distribution topology optimization. We use the same notation and definitions as in the

previous section. Our aim is to find an elastic structure with fixed supports, occupying a region within Ω with a volume not exceeding V , that exhibits the least compliance under a given, fixed static loading. Compliance of an elastic structure is defined as the work done by external forces $\ell(\mathbf{u})$, where $\mathbf{u} \in \mathcal{U}$ is the unique equilibrium displacement of the structure subject to the given load. By comparing the bilinear form (3.14) with the elastic energy density (3.8) and using that $\ell(\mathbf{u}) = a(\mathbf{u}, \mathbf{u})$ at equilibrium, we see that minimizing compliance is equivalent to minimizing the elastic energy. A very compliant material yields much when subject to load; that is, $\ell(\mathbf{u})$ is large. On the other hand, a stiff material is characterized by a small $\ell(\mathbf{u})$. Hence, compliance is an inverse measure of stiffness, thus minimizing the compliance of an elastic structure corresponds to maximizing its global stiffness.

We introduce a relaxed material indicator function $\rho \in \mathcal{D}$, where

$$\mathcal{D} = \{\rho \in L^\infty(\Omega) \mid 0 \leq \rho \leq 1 \text{ almost everywhere in } \Omega\}. \quad (3.19)$$

The spatially varying elastic properties of any design $\rho \in \mathcal{D}$ are assumed to be given by $\tilde{\rho}(\rho)\mathbf{E}$, where $\tilde{\rho}$ is termed the physical density. In order to capture the effects of the spatially varying elasticity tensor, the bilinear form in (3.14) is modified as follows,

$$a(\rho; \mathbf{u}, \mathbf{v}) = \int_{\Omega} \tilde{\rho}(\rho) \mathbf{E} \boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{v}). \quad (3.20)$$

The physical density $\tilde{\rho} : \mathcal{D} \rightarrow [\underline{\rho}, 1]$ is defined for $x \in \Omega$ by

$$\tilde{\rho}(\rho)(x) = \underline{\rho} + (1 - \underline{\rho})P(F(\rho)(x)), \quad (3.21)$$

where F is a filter operator that maps $\rho \in \mathcal{D}$ to the function $F(\rho) : \Omega \rightarrow [0, 1]$, and $P : [0, 1] \rightarrow [0, 1]$ is a smooth and strictly increasing penalty function. For instance, $P(x) = x^p$, for some $p > 1$, is the penalty function used in the popular SIMP scheme [3]. The introduction of a minimal physical density $0 < \underline{\rho} \ll 1$ implies that void is approximated by a very compliant material. This approximation greatly facilitates the analysis and solution of the topology optimization problem at hand. The idea of the penalty function is to make the contribution of intermediate values to the stiffness of the design unproportionate to their contribution to the volume constraint

$$\int_{\Omega} F(\rho) \leq V; \quad (3.22)$$

that is, intermediate values will increase the volume of the structure but not much increase the stiffness.

Since $|a(\rho; \mathbf{u}, \mathbf{v})| \leq |a(1; \mathbf{u}, \mathbf{v})| = |a(\mathbf{u}, \mathbf{v})|$, the modified bilinear form is continuous with the same constant C_1 as in (3.16). Similarly, we find that $a(\cdot; \cdot, \cdot)$ is \mathcal{U} -coercive, since

$$a(\rho; \mathbf{v}, \mathbf{v}) \geq \underline{\rho} a(\mathbf{v}, \mathbf{v}) \geq \underline{\rho} C_3 \|\mathbf{v}\|_{\mathcal{U}}^2 \text{ for all } \mathbf{v} \in \mathcal{U}, \quad (3.23)$$

where C_3 is the constant from (3.18). As in the previous section, using the Lax–Milgram lemma, we conclude that for each $\rho \in \mathcal{D}$ there is one and only one $\mathbf{u} \in \mathcal{U}$ satisfying the following variational problem.

$$\text{Find } \mathbf{u} \in \mathcal{U} \text{ such that } a(\rho; \mathbf{u}, \mathbf{v}) = \ell(\mathbf{v}) \text{ for all } \mathbf{v} \in \mathcal{U}. \quad (3.24)$$

We define $G : \mathcal{D} \rightarrow \mathcal{U}$ such that $G(\rho) = \mathbf{u}$, where \mathbf{u} is the unique solution to (3.24).

The set of feasible designs $\mathcal{A} \subset \mathcal{D}$ is defined by restricting \mathcal{D} with the volume constraint (3.22), that is,

$$\mathcal{A} = \left\{ \rho \in \mathcal{D} \mid \int_{\Omega} F(\rho) \leq V \right\}. \quad (3.25)$$

Furthermore, let \mathcal{U}^* denote the image of \mathcal{A} under G ; that is, $\mathcal{U}^* = G(\mathcal{A})$. We are now ready to state the optimization problem.

$$\text{Find } \rho^* \in \mathcal{A} \text{ and } \mathbf{u}^* = G(\rho^*) \text{ such that } \ell(\mathbf{u}^*) \leq \ell(\mathbf{u}) \text{ for all } \mathbf{u} \in \mathcal{U}^*. \quad (3.26)$$

Below follows an alternative formulation of the minimal compliance problem.

$$\text{Find } \mathbf{u}^* \in \mathcal{U}^* \text{ such that } \ell(\mathbf{u}^*) = \inf_{\mathbf{u} \in \mathcal{U}^*} \ell(\mathbf{u}). \quad (3.27)$$

Note that if ρ^* and \mathbf{u}^* solve problem (3.26), then \mathbf{u}^* solves problem (3.27). On the other hand, if \mathbf{u}^* solves problem (3.27), then there exists $\rho^* \in \mathcal{A}$ such that ρ^* and \mathbf{u}^* solve problem (3.26). Thus, existence of solutions to (3.26) implies existence of solutions to (3.27), and vice versa.

We will now present some facts regarding the existence of solutions to minimization problems (3.26) and (3.27). If both filter and penalization are removed (that is, they are replaced by identity maps) the resulting problem possesses a unique solution; in two dimensions this problem corresponds to the so-called variable thickness sheet problem [3]. The introduction of a penalty function or a discreteness constraint of the form $(1 - \rho)(\rho - \underline{\rho}) = 0$ almost everywhere in Ω in the variable thickness sheet problem leads to ill-posed problems [4]. In 2001, Bourdin [5] proved that there exists solutions to problem (3.27) when using a linear filter function, realized by taking the convolution of ρ by a positive normalized kernel with compact support; that is, the penalized and linearly filtered variable thickness sheet problem possesses at least one solution.

To solve optimization problem (3.26) numerically, with the aid of a computer, we partition the design domain into n elements. The relaxed material indicator function is then approximated by a piecewise constant function, here represented by the design vector $\rho \in [0, 1]^n$. To limit a priori bias on the optimized design, it is customary to partition the design domain by using a Cartesian grid. The use of a Cartesian grid also greatly facilitates implementation of the finite element method and the filtering procedure. A Cartesian grid is characterized by the number of elements in each coordinate direction n_i , $i \in \{1, \dots, d\}$ and the linear size of each element $h = \sqrt[d]{v}$, where v is the volume of an element in the grid. From this point and onwards, unless otherwise stated, we will consider Cartesian grids. By using a finite element method with $N + M$ nodes, the approximation of the solution to the variational formulation (3.24) satisfies

$$\mathbf{K}(\rho)\mathbf{u} = \mathbf{f}, \quad (3.28)$$

where $\mathbf{K}(\rho) \in \mathbb{R}^{N \times N}$ is the symmetric and positive definite stiffness matrix, $\mathbf{u} \in \mathbb{R}^N$ and $\mathbf{f} \in \mathbb{R}^N$ are the free nodal displacement and load vectors respectively. The remaining M nodes are located on Γ_D and are hence not free to move. The stiffness matrix can be formed by summing the contributions from all elements in the grid,

$$\mathbf{K}(\rho) = \sum_{i=1}^n \left[\underline{\rho} + (1 - \underline{\rho})P(F_i(\rho)) \right] \mathbf{K}^{(i)}, \quad (3.29)$$

where $\mathbf{K}^{(i)}$ is the element stiffness matrix corresponding to an element filled with material, and $\mathbf{F} : [0, 1]^n \rightarrow [0, 1]^n$ is the discrete version of the filter. The symmetry and positive definiteness of the stiffness matrix follows from the symmetry and \mathcal{U} -coerciveness of the bilinear form (3.20). The discrete version of the volume constraint in (3.25) is given by

$$\mathbf{1}_n^T \mathbf{F}(\boldsymbol{\rho}) \leq Vh^{-d} = nV^*, \quad (3.30)$$

where $\mathbf{1}_n = (1, \dots, 1)^T \in \mathbb{R}^n$ and V^* is the maximum allowable volume fraction. By using $\mathbf{f}^T \mathbf{u}$ as the discrete analogue of compliance, the discrete counterpart of optimization problem (3.26) reads

$$\begin{aligned} & \min_{(\boldsymbol{\rho}, \mathbf{u}) \in [0, 1]^n \times \mathbb{R}^N} \mathbf{f}^T \mathbf{u} \\ & \text{such that } \mathbf{K}(\boldsymbol{\rho}) \mathbf{u} = \mathbf{f}, \\ & \mathbf{1}_n^T \mathbf{F}(\boldsymbol{\rho}) \leq nV^*. \end{aligned} \quad (3.31)$$

As seen in (3.26), our interest lies in finding $\boldsymbol{\rho}^* \in [0, 1]^n$ such that the minimum in (3.31) is attained at $(\boldsymbol{\rho}^*, \mathbf{u}^*)$, where $\mathbf{u}^* = \mathbf{K}(\boldsymbol{\rho}^*)^{-1} \mathbf{f}$.

Since we want to use a gradient based optimization algorithm to solve (3.31), derivatives need to be evaluated. Note that we thus need to assume that the filter function \mathbf{F} is differentiable. In order to derive an expression for the objective function gradient, we introduce a design perturbation $\delta \boldsymbol{\rho}$. Since the load \mathbf{f} is design independent (compare with ℓ in (3.15)), the resulting first order perturbation of the objective function in (3.31) is given by

$$\delta(\mathbf{f}^T \mathbf{u}) = \mathbf{f}^T \delta \mathbf{u}, \quad (3.32)$$

where $\delta \mathbf{u}$ denotes the first order perturbation of the displacement vector. By using equilibrium equation (3.28) and the symmetry of $\mathbf{K}(\boldsymbol{\rho})$, we find that

$$\delta(\mathbf{f}^T \mathbf{u}) = \mathbf{u}^T \mathbf{K}(\boldsymbol{\rho}) \delta \mathbf{u}. \quad (3.33)$$

The relation between the first order perturbations of \mathbf{u} and $\mathbf{K}(\boldsymbol{\rho})$ is found by perturbing the equilibrium equation (3.28),

$$0 = \delta(\mathbf{K}(\boldsymbol{\rho}) \mathbf{u} - \mathbf{f}) = \delta \mathbf{K}(\boldsymbol{\rho}) \mathbf{u} + \mathbf{K}(\boldsymbol{\rho}) \delta \mathbf{u}. \quad (3.34)$$

Equation (3.34) and expression (3.33) yields that

$$\delta(\mathbf{f}^T \mathbf{u}) = -\mathbf{u}^T \delta \mathbf{K}(\boldsymbol{\rho}) \mathbf{u}. \quad (3.35)$$

From expression (3.29), we find that

$$\delta \mathbf{K}(\boldsymbol{\rho}) = \frac{\partial \mathbf{K}(\boldsymbol{\rho})}{\partial \rho_j} \delta \rho_j = \left(\sum_{i=1}^n \mathbf{K}^{(i)} (1 - \underline{\rho}) P'(F_i(\boldsymbol{\rho})) \frac{\partial F_i}{\partial \rho_j} \right) \delta \rho_j. \quad (3.36)$$

Expressions (3.35) and (3.36) finally results in

$$\begin{aligned} \delta(\mathbf{f}^T \mathbf{u}) &= \left(\frac{\partial}{\partial \rho_j} \mathbf{f}^T \mathbf{u} \right) \delta \rho_j = \\ &= \left(\sum_{i=1}^n (-\mathbf{u}^T \mathbf{K}^{(i)}(\boldsymbol{\rho}) \mathbf{u}) (1 - \underline{\rho}) P'(F_i(\boldsymbol{\rho})) \frac{\partial F_i}{\partial \rho_j} \right) \delta \rho_j. \end{aligned} \quad (3.37)$$

By inspection of expression (3.37), we identify the sum within the large parenthesis as the j th component of $\nabla(\mathbf{f}^T \mathbf{u})$. The gradient of the volume constraint function in (3.30) is found by direct computation.

It is natural to require that increasing the value of a design variable must not decrease any value of the physical design, that is, to require that

$$\frac{\partial F_i}{\partial \rho_j} \geq 0. \quad (3.38)$$

By combining the monotonicity (3.38) of \mathbf{F} , the assumption that $P' \geq 0$ (recall that the penalty function P is assumed to be smooth and strictly increasing), the fact that $\mathbf{v}^T \mathbf{K}^{(i)} \mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{R}^N$, and expression (3.36), we find that

$$\mathbf{v}^T \frac{\partial \mathbf{K}(\boldsymbol{\rho})}{\partial \rho_j} \mathbf{v} \geq 0 \text{ for all } \mathbf{v} \in \mathbb{R}^N. \quad (3.39)$$

The positive semidefiniteness (3.39) together with expression (3.37) implies that adding material to the design cannot increase the compliance, that is

$$\frac{\partial}{\partial \rho_j} \mathbf{f}^T \mathbf{u} \leq 0. \quad (3.40)$$

That all derivatives of the objective function are of the same sign greatly facilitates the solution of the optimization problem (3.31) by enabling the use of specialized algorithms such as the optimality criteria (OC) method [3]. However, for more complicated optimization problems, general optimization algorithms, such as the method of moving asymptotes by Svanberg [16, 17], must be used.

4. Density filters in topology optimization

Inspired by the use of noise reducing filters in image analysis, Sigmund [13] introduced the sensitivity filter in which the gradient of the objective function is filtered. The sensitivity filter is known to be effective in producing mesh-independent designs [3]. However, in this thesis, our main focus is so-called density filters where the design variables are filtered. The first density filter considered was that of Bruns and Tortorelli [6], who defined the physical design in element i as a linear weighted arithmetic average over a neighborhood with fixed (mesh-independent) radius R . Detailed comparisons of different filters for a variety of topology optimization problems are found in the papers by Sigmund [14] and Svanberg & Svård [18]. In this section, we mainly discuss filters in the discretized setting. We refer to any function $\mathbf{F} : [0, 1]^n \rightarrow [0, 1]^n$ as a general filter, and given a function $f : [0, 1] \rightarrow \mathbb{R}$, we define the vector $\mathbf{f}(\boldsymbol{\rho}) = (f(\rho_1), \dots, f(\rho_n))^T \in \mathbb{R}^n$.

A general linear filter is defined as

$$\mathbf{F}(\boldsymbol{\rho}) = \mathbf{W}\boldsymbol{\rho}, \quad (4.1)$$

where $\mathbf{W} = [w_{ij}] \in \mathbb{R}^{n \times n}$ is a normalized weight matrix with non negative entries; that is,

$$\begin{aligned} \mathbf{W}\mathbf{1}_n &= \mathbf{1}_n, \\ w_{ij} &\geq 0. \end{aligned} \quad (4.2)$$

The entries of \mathbf{W} implicitly define neighborhoods $\mathcal{N}_i \subset \{1, \dots, n\}$, $i \in \{1, \dots, n\}$ by the relation

$$w_{ij} > 0 \text{ if and only if } j \in \mathcal{N}_i. \quad (4.3)$$

We say that a collection of neighborhoods is symmetric, if for any pair of elements i, j , we have that $i \in \mathcal{N}_j$ implies that $j \in \mathcal{N}_i$. In topology optimization, it is common to use a neighborhood shape $\mathcal{N} \subset \mathbb{R}^d$ to define neighborhoods

$$\mathcal{N}_i = \{j : x_j - x_i \in \mathcal{N}\}, \quad (4.4)$$

where x_i , $i \in \{1, \dots, n\}$ are the element centroids. The predominant choice of shapes \mathcal{N} has been to use a spherical (in 3D) or circular (in 2D) neighborhood shape with radius R in combination with weights decaying linearly with the distance from the neighborhood center. Intuitively the linear filter achieves mesh-independence by imposing a minimum length scale of size about $2R$. We note that the introduction of a minimum length scale often is desirable from a manufacturing point of view.

Lazarov and Sigmund [12] introduced a rather different approach to the linear filter (4.1), where the filtered relaxed material indicator function $F(\rho)$ is the solution of the boundary value problem

$$\begin{aligned} -a^2 \Delta F(\rho) + F(\rho) &= \rho \text{ in } \Omega, \\ \frac{\partial}{\partial \mathbf{n}} F(\rho) &= 0 \text{ on } \partial\Omega, \end{aligned} \quad (4.5)$$

where $a > 0$ is a parameter similar to R . The solution of (4.5) can be written as a convolution of ρ and a positive normalized Green's function. Instead of discretizing the convolution

integral, which would lead to an expression of the form (4.1), the effect of the filter in the discretized setting is found by solving a discrete analogue of (4.5). We refer to filters defined in this way by a partial differential equation (PDE) as PDE-based filters. Since the Green's function is positive, the entries of the resulting weight matrix are all positive. According to (4.3) this means that each element has all elements in its neighborhood; that is, $\mathcal{N}_i = \{1, \dots, n\}$ for all i . Moreover, the weights w_{ij} decay with the distance from the neighborhood center x_i and the decay is increased for smaller a .

A drawback with the linear density filter (4.1) is that it tends to produce optimized designs with relatively large areas of intermediate values, hence counteracting the effect of the penalization. More recently, a whole range of *nonlinear* filters, aimed at reducing the amount of intermediate values in the final design while retaining mesh-independence, have been presented. Different filters can lead to different designs with similar performance. It has been argued that using a variety of filters gives the designer freedom to decide based on other preferences than those expressed by the objective function and constraints [18].

The Heaviside filter [9] consists of applying a linear filter followed by the application of a smooth approximation of the Heaviside step function,

$$\mathbf{F}(\boldsymbol{\rho}) = \mathbf{1}_n - e^{-\beta \mathbf{W}\boldsymbol{\rho}} + e^{-\beta} \mathbf{W}\boldsymbol{\rho}, \quad (4.6)$$

where $\beta \in [0, \infty)$ is a parameter controlling the sharpness of the step function approximation. For $\beta = 0$ the linear filter (4.1) is retrieved, and when $\beta \rightarrow \infty$ the Heaviside filter approaches the discontinuous filter

$$F_i(\boldsymbol{\rho}) = \begin{cases} 0 & \text{if } \rho_j = 0 \text{ for all } j \in \mathcal{N}_i, \\ 1 & \text{otherwise.} \end{cases} \quad (4.7)$$

We note that the step in the approximation of the Heaviside step function is located at 0. By using a different approximation of the Heaviside step function, with the step located at $\eta \in [0, 1]$, Wang et al. [19] presented the following variation of the Heaviside filter,

$$\mathbf{F}(\boldsymbol{\rho}) = \frac{\tanh(\beta\eta)\mathbf{1}_n + \tanh(\beta(\mathbf{W}\boldsymbol{\rho} - \eta\mathbf{1}_n))}{\tanh(\beta\eta) + \tanh(\beta(1 - \eta))}, \quad (4.8)$$

where as before $\beta \in (0, \infty)$ is a parameter controlling the sharpness of the step function approximation. In the limit of vanishing β , for a fixed η , the filter in (4.8) approaches the linear filter (4.1), while for $\beta \rightarrow \infty$ and a fixed $\eta \in (0, 1)$ the filter approaches the discontinuous filter

$$F_i(\boldsymbol{\rho}) = \begin{cases} 0 & \text{if } (\mathbf{W}\boldsymbol{\rho})_i < \eta, \\ 1/2 & \text{if } (\mathbf{W}\boldsymbol{\rho})_i = \eta, \\ 1 & \text{if } (\mathbf{W}\boldsymbol{\rho})_i > \eta. \end{cases} \quad (4.9)$$

The filters (4.6) and (4.8) has both been categorized as projection based filters [15], where the term projection refers to their capability of "projecting" $\boldsymbol{\rho} \in [0, 1]^n$ onto $\mathbf{F}(\boldsymbol{\rho}) \in \{0, 1\}^n$ in the limit of large β . (As seen in (4.9), in the limit of large β for a fixed $\eta \in (0, 1)$ the filter (4.8) would actually produce physical designs with values in $\{0, 1/2, 1\}$. However, the exceptional value 1/2 is rather improbable.) For the same reason the application of the approximate Heaviside step function is sometimes referred to as a projection step. Technically, even in the limit of large β , the filters (4.6) and (4.8) are not projections since they fail to be idempotent; that is, in general $\mathbf{F}(\mathbf{F}(\boldsymbol{\rho})) \neq \mathbf{F}(\boldsymbol{\rho})$.

In 2007 Sigmund [14] introduced filters based on morphology operators [11] from image analysis. These morphology-based filters are not generated by adding a projection step to the linear filter (4.1). The dilate filter [14], which is based on the exponential average, is given by

$$\mathbf{F}(\boldsymbol{\rho}) = \frac{1}{\beta} \ln (\mathbf{W} \mathbf{e}^{\beta \boldsymbol{\rho}}), \quad (4.10)$$

where $\beta \in (0, \infty)$ is a parameter controlling the properties of the filter. In the limit of vanishing β , the dilate filter (4.10) approaches the linear filter (4.1), while for $\beta \rightarrow \infty$, the dilate filter approaches the discontinuous moving maximum filter

$$F_i(\boldsymbol{\rho}) = \max_{j \in \mathcal{N}_i} \rho_j. \quad (4.11)$$

The term dilate relates to the fact that on a binary design, the moving maximum filter (4.11) dilates the regions corresponding to material. More precisely, if element i is occupied with material then the moving maximum filter fills all elements j that have i in their neighborhood with material. We remark that if the collection of neighborhoods is symmetric, then this amounts to filling the entire neighborhood \mathcal{N}_i with material.

A means to arrive at new filters is to use the flip transformation $\boldsymbol{\rho} \rightarrow \mathbf{1}_n - \boldsymbol{\rho}$, which interchanges the roles of material and void [18]. Applying the flip transformation to the dilate filter (4.10) leads to the erode filter [14]

$$\mathbf{F}(\boldsymbol{\rho}) = \mathbf{1}_n - \frac{1}{\beta} \ln (\mathbf{W} \mathbf{e}^{\beta (\mathbf{1}_n - \boldsymbol{\rho})}) = -\frac{1}{\beta} \ln (\mathbf{W} \mathbf{e}^{-\beta \boldsymbol{\rho}}), \quad (4.12)$$

where we have used the normalization of \mathbf{W} to reach the rightmost form. As for the dilate filter (4.10), the erode filter (4.12) approaches the linear filter (4.1) in the limit of vanishing β . However, when $\beta \rightarrow \infty$ the erode filter approaches the discontinuous moving minimum filter

$$F_i(\boldsymbol{\rho}) = \min_{j \in \mathcal{N}_i} \rho_j. \quad (4.13)$$

Since the moving minimum filter can be obtained by using the flip transformation on the moving maximum filter (4.11), it acts by dilating the void regions, which is the same as eroding the material regions. More precisely, if element i is empty (occupied with void) the moving minimum filter empties all elements j that have i in their neighborhood. We remark that if the collection of neighborhoods is symmetric, then this amounts to removing all material within the neighborhood \mathcal{N}_i .

Svanberg and Svård [18] introduced erode-like filters based on the remaining two Pythagorean means¹, the harmonic erode filter

$$\mathbf{F}(\boldsymbol{\rho}) = (\mathbf{W}(\boldsymbol{\rho} + \alpha \mathbf{1}_n)^{-1})^{-1} - \alpha \mathbf{1}_n, \quad (4.14)$$

and the geometric erode filter

$$\mathbf{F}(\boldsymbol{\rho}) = \mathbf{e}^{\mathbf{W} \ln(\boldsymbol{\rho} + \alpha \mathbf{1}_n)} - \alpha \mathbf{1}_n, \quad (4.15)$$

respectively, where the parameter $\alpha \in (0, \infty)$ controls the behavior of the filters. For $\alpha \rightarrow \infty$ both the harmonic (4.14) and the geometric erode filter (4.15) approach the linear

¹The Pythagorean means consists of the arithmetic, the harmonic and the geometric mean.

filter (4.1). These filters are in general not approaching the moving minimum filter when $\alpha \rightarrow 0$. However, for any $\boldsymbol{\rho} \in [0, 1]^n$ such that $\min_{j \in \mathcal{N}_i} \rho_j = 0$, we find that $F_i(\boldsymbol{\rho})$ approaches $0 = \min_{j \in \mathcal{N}_i} \rho_j$ as $\alpha \rightarrow 0$. Dilate-like filters are obtained from the erode-like filters (4.14) and (4.15) by using the flip transformation [18].

The bilateral density filter [20] takes a form that is rather different from the filters presented so far, namely

$$\mathbf{F}(\boldsymbol{\rho}) = \frac{\hat{\mathbf{W}}(\boldsymbol{\rho})\boldsymbol{\rho}}{\hat{\mathbf{W}}(\boldsymbol{\rho})\mathbf{1}_n}, \quad (4.16)$$

where $\hat{\mathbf{W}}(\boldsymbol{\rho}) = \mathbf{W} \cdot [\tilde{w}(|\rho_i - \rho_j|)] \in \mathbb{R}^{n \times n}$, \cdot denotes the element wise product, and $\tilde{w} : [0, 1] \rightarrow (0, \infty)$ is a non increasing function. The value of the bilateral density filter (4.16) at element i consists of a nonlinearly weighted average over the neighborhood \mathcal{N}_i , where the weights not only depend on the physical location of the elements (w_{ij}) but also on the deviation of the values from ρ_i ($\tilde{w}(|\rho_i - \rho_j|)$).

All the filters (4.10) to (4.16) could be termed internal since they respect the actual range of $\boldsymbol{\rho}$,

$$\min_{j \in \mathcal{N}_i} \rho_j \leq F_i(\boldsymbol{\rho}) \leq \max_{j \in \mathcal{N}_i} \rho_j. \quad (4.17)$$

We remark that the projection based filters (4.6) and (4.8) in general violate condition (4.17).

One way of constructing filters with new properties is to apply different filters in sequence; for instance by combining a dilate (**D**) and an erode (**E**) filter we arrive at either the close (**ED**) or the open (**DE**) filter depending on the order of application [14]. In the limit of maximum and minimum operators, the close filter removes holes (void regions) smaller than the filter size, while the open filter removes structural elements (material regions) smaller than the filter size. Hence, the open filter provides minimum length scale control on structural elements, while the close filter provides minimum length scale control on void regions. Combining the close and open filters results in the close-open (**DEED**) or open-close (**EDDE**) filters that in some cases provide minimum length scale control on both material and void regions [14].

We see that many filters used in topology optimization are based on weighted averages over some neighborhoods. Computation of the average over \mathcal{N}_i requires $O(|\mathcal{N}_i|)$ arithmetic operations; that is, the computational complexity grows linearly with the size of the neighborhood. If neighborhoods with fixed physical sizes are used we have that $|\mathcal{N}_i| \propto n$ for all i , which shows that the computational complexity of computing averages over all neighborhoods can be estimated as $O(n^2)$. By using a properly tuned multigrid conjugate gradient method the linear system (3.28) can be solved in $O(n)$ operations [1]. Hence, although the leading coefficient ($\sim |\mathcal{N}_i|/n$) in the $O(n^2)$ estimate is rather small in practice, the computational cost for filtering can be substantial for large scale problems. Some optimization algorithms, such as the OC method mentioned at the end of the previous section, relies on multiple evaluations of the volume constraint (3.30) for computing each design update. The computational complexity of such algorithms are in general dependent on the computational complexity of the filtering procedure. By using the PDE approach, described in conjunction with boundary value problem (4.5), together with a properly tuned multigrid method, linearly weighted arithmetic averages over all neighborhoods can be computed in $O(n)$ operations. However, in this case, we are not free to choose the weights since \mathbf{W} is dictated by the PDE.

When using a gradient based optimization method, the computational cost of evaluating gradients must also be considered. In particular we are interested in determining the extra

cost related to the filter. To this end; let J be a function that depends on $\boldsymbol{\rho}$ via $\mathbf{F}(\boldsymbol{\rho})$, then the components of ∇J can be found by using the chain rule,

$$\frac{\partial J}{\partial \rho_j} = \frac{\partial J}{\partial F_i} \frac{\partial F_i}{\partial \rho_j}. \quad (4.18)$$

In the absence of a filter; that is, if $\mathbf{F}(\boldsymbol{\rho}) = \boldsymbol{\rho}$ for all $\boldsymbol{\rho}$, then $\partial J / \partial F_j$ would be the j th component of ∇J . Therefore, expression (4.18) may be interpreted as a modification or even as a filtering (compare with (4.1)) of the unfiltered gradient of J . Furthermore, expression (4.18) shows that the additional cost related to the filter can be found by considering $\boldsymbol{\xi}^T \nabla \mathbf{F}$ for some $\boldsymbol{\xi} \in \mathbb{R}^n$.

5. Thesis contributions

In this section, we summarize some of the thesis contributions.

5.1 General filters and the class of fW -mean filters

In paper II, we present some typical requirements on a general filter function $\mathbf{F} : [0, 1]^n \rightarrow [0, 1]^n$. The filter should be increasing in the sense that for any pair i, j and any $\delta \geq 0$,

$$F_i(\boldsymbol{\rho} + \delta \mathbf{e}_j) \geq F_i(\boldsymbol{\rho}), \quad (5.1)$$

where \mathbf{e}_j denotes the j th basis vector of \mathbb{R}^n . Condition (5.1) stems from the idea that increasing a design variable should not decrease any value in the physical design. We show that, if binary physical designs are to be attainable, a consequence of the monotonicity condition (5.1) is that we are forced to require

$$\begin{aligned} \mathbf{F}(\mathbf{0}_n) &= \mathbf{0}_n, \\ \mathbf{F}(\mathbf{1}_n) &= \mathbf{1}_n, \end{aligned} \quad (5.2)$$

where $\mathbf{0}_n = (0, \dots, 0)^T \in \mathbb{R}^n$. It is also natural to require that the physical design $\mathbf{F}(\boldsymbol{\rho})$ is sensitive to changes in the design vector $\boldsymbol{\rho}$. A rather weak assumption to guarantee sensitivity to design changes, is to require that for each $\boldsymbol{\rho} \in [0, 1]^n$ there exists an i and a j such that $F_i(\boldsymbol{\rho} + \delta \mathbf{e}_j) > F_i(\boldsymbol{\rho})$ for any sufficiently small positive δ ; that is, there exists $\epsilon > 0$ such that

$$F_i(\boldsymbol{\rho} + \delta \mathbf{e}_j) > F_i(\boldsymbol{\rho}) \text{ for all } 0 < \delta < \epsilon. \quad (5.3)$$

In paper I, we introduce the class of fW -mean filters, based on the quasi-arithmetic mean, also known as the generalized f -mean. With the notation of the previous section, the fW -mean filters are of the form

$$\mathbf{F}(\boldsymbol{\rho}) = \mathbf{f}^{-1}(\mathbf{W}\mathbf{f}(\boldsymbol{\rho})), \quad (5.4)$$

where $f : [0, 1] \rightarrow \mathbb{R}$ is a smooth and invertible function with nonzero derivative, and $\mathbf{W} \in \mathbb{R}^{n \times n}$ is a weight matrix satisfying (4.2). The entries of the Jacobian of the filter (5.4) are given by

$$\frac{\partial F_i}{\partial \rho_j} = w_{ij} \frac{f'(\rho_j)}{f'(F_i(\boldsymbol{\rho}))}. \quad (5.5)$$

Note that, since i and j are free indices, summation is *not* implied in expression (5.5). By using that a continuous function defined on an interval is invertible if and only if it is strictly monotone, that $f' \neq 0$ by assumption, and that $w_{ij} \geq 0$, we find that the fW -mean filters (5.4) are increasing. That is,

$$\frac{\partial F_i}{\partial \rho_j} \geq 0, \quad (5.6)$$

where equality holds if and only if $w_{ij} = 0$. Furthermore, since \mathbf{W} is normalized, the fW -mean filters have the property of mapping a vector with equal entries to itself; that is, for any $c \in [0, 1]$,

$$\mathbf{F}(c\mathbf{1}_n) = c\mathbf{1}_n. \quad (5.7)$$

The above identity implies that the fW -mean filters satisfy (5.2). Since the properties (4.2) of \mathbf{W} imply that $\mathbf{W}\mathbf{f}(\boldsymbol{\rho}) \in [f_{\min}, f_{\max}]^n$, we find that the fW -mean filters are internal; that is, they satisfy (4.17). The class of fW -mean filters contains the linear filter (4.1), the dilate (4.10) and erode (4.12) filters, as well as the Pythagorean filters (4.14) and (4.15). To also incorporate the projection based filters (4.6) and (4.8), which are not fW -mean filters, we introduce the class of *generalized fW -mean filters*, of the form

$$\mathbf{F}(\boldsymbol{\rho}) = \mathbf{g}(\mathbf{W}\mathbf{f}(\boldsymbol{\rho})), \quad (5.8)$$

where $g : f([0, 1]) \rightarrow [0, 1]$ is a smooth function with $g(f(0)) = 0$ and $g(f(1)) = 1$. We note that definition (5.8) makes sense even if f is not a bijection. In paper I, we give a sufficient condition on the filter function (5.8) to guarantee the invertibility of f . However, the requirements on a general filter given above are too weak to force f to be invertible. Nevertheless, if f is invertible, then by defining $h = g \circ f$, we find that the generalized fW -mean filter (5.8) is nothing but a projected fW -mean filter; that is,

$$\mathbf{F}(\boldsymbol{\rho}) = \mathbf{h}(\mathbf{f}^{-1}(\mathbf{W}\mathbf{f}(\boldsymbol{\rho}))). \quad (5.9)$$

In paper I, we present an algorithm that applies fW -mean filters with linear computational complexity. The prerequisites for the algorithm are that the design variables are located in a regular grid, that neighboring elements are weighted equally, and that the neighborhood shape is a polytope. With equal weighting within a neighborhood, we can compute the average over the neighborhood by first summing and then dividing by the number of elements,

$$w_{ij}f(\rho_j) = \sum_{j \in \mathcal{N}_i} \frac{1}{|\mathcal{N}_i|} f(\rho_j) = \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} f(\rho_j). \quad (5.10)$$

The key to achieve linear computational complexity is to use a recursive update strategy for computing the neighborhood sums instead of computing each sum from scratch. More precisely, given the sum over \mathcal{N}_i , we can determine the sum over an adjacent neighborhood \mathcal{N}_j by updating, that is,

$$\sum_{k \in \mathcal{N}_j} = \sum_{k \in \mathcal{N}_i} + \sum_{k \in \mathcal{N}_j \setminus \mathcal{N}_i} - \sum_{k \in \mathcal{N}_i \setminus \mathcal{N}_j}. \quad (5.11)$$

Furthermore, since the facets of a polytope in \mathbb{R}^d are polytopes in \mathbb{R}^{d-1} , the update can be performed in a recursive manner. We show that the computational complexity of the presented summation algorithm can be bounded independent of the physical size of the neighborhoods. However, the computational complexity grows with the complexity of the neighborhood polytope. We present in detail the summation algorithm for two special cases, octagonal and rhombicuboctahedron shaped neighborhoods in 2D and 3D, respectively, and determine their computational complexity to leading order. The sizes of the octagon and rhombicuboctahedron are controlled by a parameter r , called the

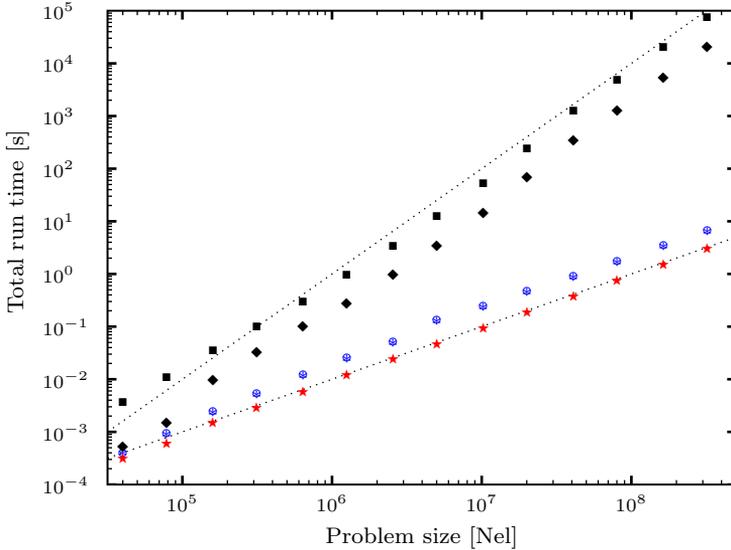


Figure 1: Run times required to compute the neighborhood sums on square domains of size n with octagonal neighborhoods of size $r = \sqrt{n}/20$ and $r = \sqrt{n}/40$ for algorithms A1 (black squares and diamonds) and A2 (blue asterisks and circles). The red stars mark the run times for algorithm A0 and the dotted lines illustrate the expected scaling.

filter radius ($|\mathcal{N}_i| \propto r^d$). To assess the computational complexity of the fast summation algorithm for octagonal shaped neighborhoods in 2D, we compared the runtimes of three different algorithms,

- A0 creating and writing n random numbers to memory using Matlab's rand function,
- A1 explicitly summing over each octagonal neighborhood, that is the ordinary $O(n^2)$ algorithm,
- A2 computing the neighborhood sums using the $O(n)$ algorithm.

The algorithm A0 essentially gives a lower bound on the time for computing the neighborhood sums and serves as a reference. To indicate the effect of the neighborhood size on the computational complexity, two different neighborhood sizes were used. In the numerical experiments, we use square domains with n elements and set the filter radius to $r = \sqrt{n}/20$ and $r = \sqrt{n}/40$, respectively. Figure 1 shows the average runtime, required for algorithms A0–A2, as a function of the number of elements in the grid. As is expected from the operation count, the runtime of algorithm A2 is essentially independent of the neighborhood size, while the runtime of algorithm A1 grows with the neighborhood size. In fact, the runtime of A1 is expected to grow linearly with the size of the neighborhood.

To enable the use of gradient based optimization algorithms, the derivatives of the objective and constraint functions must be evaluated. In paper II, we show that gradient modification related to the filter (that is, to compute $\xi^T \nabla \mathbf{F}$ for some $\xi \in \mathbb{R}^n$) can be performed with linear computational complexity by using the same algorithm as for filtering.

A cascade of generalized fW -mean filters,

$$\begin{aligned} \mathbf{C}^{(K)} &= \mathbf{F}^{(K)} \circ \mathbf{F}^{(K-1)} \circ \dots \circ \mathbf{F}^{(1)}, \\ \mathbf{F}^{(k)}(\boldsymbol{\rho}) &= \mathbf{g}_k(\mathbf{W}^{(k)} \mathbf{f}_k(\boldsymbol{\rho})) \text{ for all } k \in \{1, \dots, K\}, \end{aligned} \quad (5.12)$$

was analyzed, and we present in detail how modification of gradients can be performed with $O(Kn)$ arithmetic operations using the same algorithm as for filtering. The open, close, open–close, and close–open filters introduced, in the context of topology optimization, by Sigmund [14] are examples of cascaded fW -mean filters.

A consequence of the update strategy used to compute the neighborhood sums is that the number of elements that affect the value of a particular neighborhood sum is increased as compared to summing over each neighborhood separately. From a numerical point of view, the update strategy is potentially sensitive to round-off errors due to cancellation effects. In paper I, we present a numerical example illustrating how the element-wise accuracy might be affected by using the update strategy to compute the neighborhood sums. To minimize such numerical problems some care is needed when choosing the functions f and g .

5.2 fW -mean filtered compliance minimization

In paper II, we prove that there exists solutions to optimization problem (3.27) when using a continuous version of the fW -mean filter. For a given smooth and invertible function $f : [0, 1] \rightarrow [f_{\min}, f_{\max}] \subset \mathbb{R}$ and for any $\rho \in \mathcal{D}$, the continuous fW -mean filter $F(\rho) : x \in \Omega \rightarrow F(\rho)(x) \in [0, 1]$ is given by

$$F(\rho)(x) = f^{-1} \left(\int_{\Omega} w(x, y) (f \circ \rho)(y) \, dy \right), \quad (5.13)$$

where $w(x, \cdot) \in L^1(\Omega)$ is a non negative normalized weight function; that is, $w(x, \cdot) \geq 0$ almost everywhere in Ω , and

$$\int_{\Omega} w(x, y) \, dy = 1 \text{ for all } x \in \Omega. \quad (5.14)$$

By analogous arguments, one can show that the existence of solutions to (3.27) also follows if an extra projection step is included, or a cascade of projected fW -mean filters is used.

We have performed numerical optimization using the 2D multigrid-CG topology optimization code by Amir et al. [1] in combination with our implementation of the fast summation algorithm for octagonal shaped neighborhoods introduced in paper I. Figure 2 shows the setup used for optimizing a 2D cantilever beam that is clamped along its left end and subject to a uniformly distributed load acting on the middle 10% of its right end. We used $V^* = 0.5$ in the volume constraint (3.30), and the design domain had length to height ratio 1.5. Figure 3 presents the physical design of an cantilever beam optimized using 2160×1440 elements. We used a harmonic open–close filter that is defined analogously as the open–close filter [14] but using harmonic averaging [18] instead of exponential averaging. More precisely, the harmonic open–close filter is a cascade (5.12) of four fW -mean

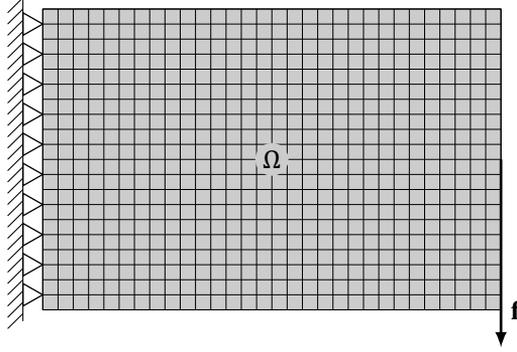


Figure 2: Problem setup for optimization of a 2D cantilever beam.

filters,

$$\begin{aligned}
 f_1(x) &= f_4(x) = (x + \alpha)^{-1}, \\
 f_2(x) &= f_3(x) = f_1(1 - x), \\
 g_k &= f_k^{-1},
 \end{aligned} \tag{5.15}$$

with the fixed parameter $\alpha = 10^{-4}$ (compare with the harmonic erode filter (4.14)). The optimized cantilever beam in figure 3 was obtained after 91 iterations and has $M_{nd} = 0.26\%$, where

$$M_{nd} = \frac{\mathbf{F}(\boldsymbol{\rho})^T (\mathbf{1}_n - \mathbf{F}(\boldsymbol{\rho}))}{n/4} \in [0, 1], \tag{5.16}$$

is the measure of non discreteness introduced by Sigmund [14] to quantify the discreteness of a physical design. A series of optimizations using different neighborhood sizes indicate that the harmonic open–close filter provides mesh-independence and independent minimum size control on material and void regions.

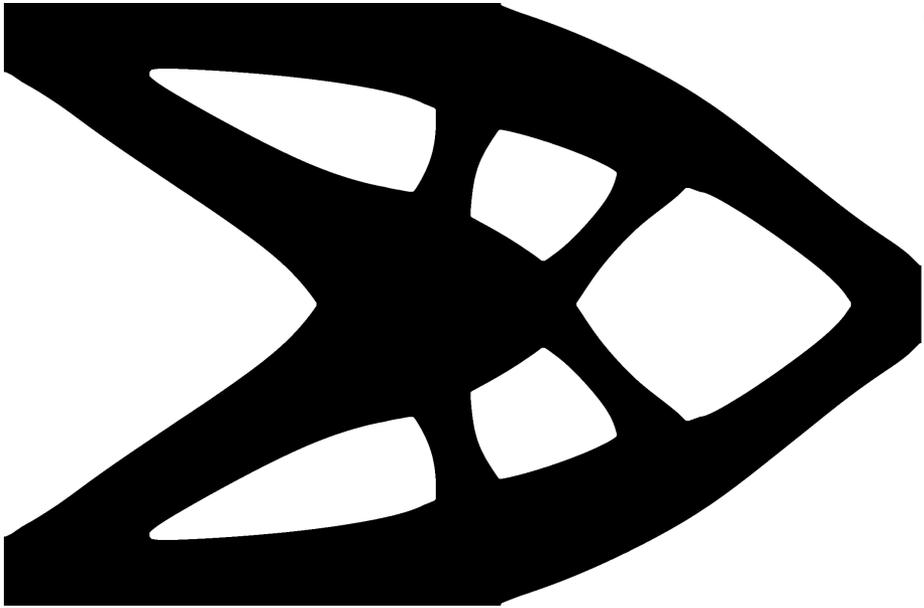


Figure 3: Physical design for an optimized cantilever beam using 2160×1440 elements, and a harmonic open filter over a “large” octagonal neighborhood followed by a harmonic close filter over a “small” octagonal neighborhood. The neighborhoods are indicated in the upper-right corner.

Bibliography

- [1] O. Amir, N. Aage, and B. S. Lazarov. On multigrid-CG for efficient topology optimization. *Structural and Multidisciplinary Optimization*, 49(5):815–829, 2014.
- [2] M. P. Bendsøe and N. Kikuchi. Generating optimal topologies in structural design using a homogenization method. *Computer Methods in Applied Mechanics and Engineering*, 71(2):197–224, 1988.
- [3] M. P. Bendsøe and O. Sigmund. *Topology Optimization. Theory, Methods, and Applications*. Springer, 2003.
- [4] T. Borrvall. Topology optimization of elastic continua using restriction. *Archives of Computational Methods in Engineering*, 8(4):351–385, 2001.
- [5] B. Bourdin. Filters in topology optimization. *International Journal for Numerical Methods in Engineering*, 50:2143–2158, 2001.
- [6] T. E. Bruns and D. A. Tortorelli. Topology optimization of non-linear elastic structures and compliant mechanisms. *Computer Methods in Applied Mechanics and Engineering*, 190:3443–3459, 2001.
- [7] P. G. Ciarlet. *Linear and nonlinear functional analysis with applications*. SIAM, 2013.
- [8] J. D. Deaton and R. V. Grandhi. A survey of structural and multidisciplinary continuum topology optimization: post 2000. *Structural and Multidisciplinary Optimization*, 49(1):1–38, 2014.
- [9] J. K. Guest, J. H. Provost, and T. Belytschko. Achieving minimum length scale in topology optimization using nodal design variables and projection functions. *International Journal for Numerical Methods in Engineering*, 61(2):238–254, 2004.
- [10] M. E. Gurtin. *An introduction to continuum mechanics*. Academic Press, 2003.
- [11] H. J. A. M. Heijmans. Mathematical morphology: A modern approach in image processing based on algebra and geometry. *SIAM Review*, 37(1):1–36, 1995.
- [12] B. S. Lazarov and O. Sigmund. Filters in topology optimization based on Helmholtz-type differential equations. *International Journal for Numerical Methods in Engineering*, 86(6):765–781, 2011.
- [13] O. Sigmund. *Design of Material Structures Using Topology Optimization*. PhD thesis, Technical University of Denmark, 1994.
- [14] O. Sigmund. Morphology-based black and white filters for topology optimization. *Structural and Multidisciplinary Optimization*, 33(4–5):401–424, 2007.

- [15] O. Sigmund and A. Maute. Topology optimization approaches. *Structural and Multidisciplinary Optimization*, 48(6):1031–1055, 2013.
- [16] K. Svanberg. The method of moving asymptotes—a new method for structural optimization. *International Journal for Numerical Methods in Engineering*, 24(2):359–373, 1987.
- [17] K. Svanberg. A class of globally convergent optimization methods based on conservative convex separable approximations. *SIAM Journal on Optimization*, 12(2):555–573, 2002.
- [18] K. Svanberg and H. Svärd. Density filters for topology optimization based on the Pythagorean means. *Structural and Multidisciplinary Optimization*, 48(5):859–875, 2013.
- [19] F. Wang, B. S. Lazarov, and O. Sigmund. On projection methods, convergence and robust formulations in topology optimization. *Structural and Multidisciplinary Optimization*, 43(6):767–784, 2011.
- [20] M. Y. Wang and S. Wang. Bilateral filtering for structural topology optimization. *International Journal for Numerical Methods in Engineering*, 63(13):1911–1938, 2005.