Optimality criteria method for optimal design problems

Crnjac, Ivana

University of Zagreb
Faculty of Science
Department of Mathematics

Ivana Crnjac

OPTIMALITY CRITERIA METHOD FOR OPTIMAL DESIGN PROBLEMS

DOCTORAL THESIS

Zagreb, 2019

University of Zagreb

Faculty of Science
Department of Mathematics

Ivana Crnjac

OPTIMALITY CRITERIA METHOD FOR OPTIMAL DESIGN PROBLEMS

DOCTORAL THESIS

Supervisors:
Assoc. Prof. Marko Vrdoljak
Assoc. Prof. Krešimir Burazin

Zagreb, 2019

Sveučilište u Zagrebu

Prirodoslovno - matematički fakultet
Matematički odsjek

Ivana Crnjac

METODA UVJETA OPTIMALNOSTI ZA ZADAĆE OPTIMALNOG DIZAJNA

DOKTORSKI RAD

Mentori:
Izv. prof. dr. sc. Marko Vrdoljak
Izv. prof. dr. sc. Krešimir Burazin

Zagreb, 2019.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisors, Prof. Krešimir Burazin and Prof. Marko Vrdoljak for the continuous support, patience, motivation, and immense knowledge that they have shared during my PhD study. Their guidance has helped me throughout my scientific research and writing of this thesis.

Many thanks to our working group in Zagreb who listened carefully all my presentations, especially to Nenad for many insightful comments and encouragement. I would also like to thank all my colleagues and friends who have always been a major source of support.

Last but not the least, I would like to thank my family: my husband Filip, my parents, Ivica and Stankica, and my brother Matija for their unconditional love and support during my scientific jurney and my life in general.

Prije svega, želim izraziti iskrenu zahvalnost svojim mentorima, prof. Krešimiru Burazinu i prof. Marku Vrdoljaku na neprekidnoj podršci, strpljenju, motivaciji i neizmjernom znanju koje su podijelili sa mnom tijekom studiranja. Njihovo vođenje uvelike mi je pomoglo tijekom znanstvenog istraživanja i pisanja ove teze.

Zahvaljujem i našoj radnoj grupi u Zagrebu koja je pažljivo slušala sva moja izlaganja, posebno Nenadu na mnogim pronicljivim komentarima i poticanju. Također, hvala svim mojim kolegama i prijateljima koji su uvijek bili veliki izvor podrške.

Posljednje, ali ne i najmanje važno, želim se zahvaliti svojoj obitelji: suprugu Filipu, roditeljima Ivici i Stankici, te bratu Matiji na njihovoj bezuvjetnoj ljubavi i podršci na mom znanstvenom putu te životu općenito.

Summary

In this thesis, we study numerical solutions for optimal design problems. In such problems, the goal is to find an arrangement of given materials within the domain which minimizes (or maximizes) a particular integral functional, under constraints on the amount of materials and PDE constraints that underlay involved physics. We consider such problems in the frame of the stationary diffusion equation and linearized elasticity system for domains occupied by two isotropic materials.

In Chapter 1 we review the basic facts about homogenization theory. The definition of H-convergence and composite materials is presented as well as some of their main properties.

Chapter 2 focuses on the multiple state optimal design problems for the stationary diffusion equation. We give necessary condition of optimality for the relaxed formulation of optimal design problems, where the relaxation was obtained by the homogenization method. Moreover, we present a new variant of the optimality criteria method suitable for some minimization problems. The method is tested on various examples, and convergence is proved in the spherically symmetric case and the case when the number of states is less then the space dimension.

The single state optimal design problem in linearized elasticity is addressed in Chapter 3. We give an explicit calculation of the lower Hashin-Shtrikman bound on the complementary energy in two and three space dimensions, and derive the optimality criteria method for two-dimensional compliance minimization problems.

Keywords: Optimality criteria method, multiple state optimal design problems, homogenization, stationary diffusion, linearized elasticity, convergence, Hashin-Shtrikman bounds;

Sažetak

Zadaće optimalnog dizajna pojavljuju se u raznim područjima fizike, mehanike, arhitekture, medicine i sl. Primjerice, minimizacija ukupne količine topline u tijelu ili maksimizacija protoka dva viskozna nemješiva fluida kroz cijev tipični su problemi optimalnog dizajna. U zadaćama optimalnog dizajna cilj je pronaći raspodjelu danih materijala, tako da dobiveno tijelo zadovoljava neke kriterije optimalnosti. Optimalnost raspodjele obično se izražava kroz minimizaciju (maksimizaciju) određenog integralnog funkcionala, uz uvjete na količinu materijala i danu parcijalnu diferencijalnu jednadžbu koja opisuje makroskopska obilježja proučavanog problema.

S obzirom da klasična rješenja (dizajni) najčešće ne postoje, polaznu zadaću je potrebno relaksirati. U prvom poglavlju dajemo pregled osnovnih rezultata teorije homogenizacije, koju koristimo za relaksaciju polaznog problema. Skup svih generaliziranih dizajna, poznat pod nazivom G-zatvarač, poznat je u slučaju mješavine dvaju izotropnih materijala s aspekta vodljivosti, te se disertacija fokusira upravo na dvofazni optimalni dizajn. Zbog svoje kompleksnosti, probleme optimalnog dizajna u praksi je moguće rješavati uglavnom numeričkim metodama, te se u sljedećim poglavljima daju novi numerički algoritmi za rješavanje ovih zadaća.

Drugo poglavlje bavi se problemima optimalnog dizajna u vodljivosti s više jednadžbi stanja. Koristeći svojstva G-zatvarača, računamo nužne uvjete optimalnosti za relaksirani problem, koje potom koristimo u razvoju nove varijante metode uvjeta optimalnosti. Pokazuje se da dobiveni algoritam daje konvergentni niz dizajna za neke minimizacijske probleme. Štoviše, u disertaciji je dan dokaz konvergencije algoritma u slučaju kada je broj jednadžbi stanja manji od dimenzije domene te u sferno simetričnom slučaju, gdje se optimalni dizajn može pronaći među jednostavnim laminama.

Za probleme linearizirane elastičnosti G-zatvarač nije poznat čak ni za mješavine dvaju izotropnih materijala. No, u posebnim slučajevima korisnim se pokazuju tzv. Hashin-Shtrikmanove ocjene na skup svih mogućih mješavina. Prema tome, u posljednjem poglavlju bavimo se minimizacijom potencijalne elastične energije s jednom jednadžbom stanja. Kako se u nužnom uvjetu optimalnosti za ovaj problem javlja donja Hashin-Shtrikmanova ocjena na komplementarnu energiju, eksplicitno ju računamo u dvodimenzionalnom slučaju, a potom izračunato koristimo u razvoju nove varijante metode uvjeta optimalnosti. Također, dajemo eksplicitnu Hashin-Shtrikmanovu ocjenu u trodimenzion-

alnom slučaju.

Ključne riječi: Metoda uvjeta optimalnosti, optimalni dizajn s više jednadžbi stanja, homogenizacija, jednadžba stacionarne difuzije, linearizirana elastičnost, konvergencija, Hashin-Shtrikmanove ocjene;

Contents

Introduction

Optimal design is a broad field of research in applied mathematics, since it arises in various areas like physics, mechanics, architecture, medicine, etc. For example, minimization of the amount of heat kept inside the body, or maximization the flow rate of two viscous immiscible fluids through pipe are typical problems of optimal design. In optimal design problems, the goal is to find the best arrangement of given materials within the body, which optimizes its properties with respect to some optimality criteria. Optimality of the distribution is usually expressed as a minimization (maximization) of an integral functional of form

$$I(\boldsymbol{\chi}) = \int_{\Omega} F(\mathbf{x}, u(\mathbf{x}), \boldsymbol{\chi}(\mathbf{x}))\, d\mathbf{x},$$

depending on rearrangement $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_k) \in \mathrm{L}^\infty(\Omega; K)$, $K = \{\boldsymbol{\kappa} \in \{0,1\}^k : \sum_{j=1}^k \kappa_j = 1\}$, of materials that constitute the domain and solution u of a partial differential equation modelling the involved physics. Optimal design can be seen as a special branch of optimal control problems, where the function $\boldsymbol{\chi}$ acts as a control.

These problems usually do not admit classical solutions (designs), so an appropriate relaxation of the original problem is needed. It consists of finding an adequate space of admissible designs in which the problem is well-posed. Murat and Tartar showed that the homogenization method, where a mixture of original materials on micro scale is used as a generalized design, gives a proper relaxation of the original problem (Murat & Tartar (1985)). Here, the question of characterizing the set of all possible mixtures obtained by the homogenization process arises, known as the G-closure problem. The set is known in the case of mixing two isotropic materials from the conductivity point of view (Tartar (1985), Lurie & Cherkaev (1984)), while in the linearized elasticity setting this is still an open question, even in the case of mixing two isotropic materials. In this case, Hashin-Shtrikman bounds on the set of all possible mixtures reveal useful (Francfort & Murat (1986), Tartar (1986), Zhikov et al. (1994)). The dissertation will be focused on mixtures of two isotropic materials in the conductivity and linearized elasticity setting.

In the case of single stationary diffusion equation with a constant (heat, charge) source term, Murat and Tartar proved (Murat & Tartar (1985)) that the problem of maximization of the energy functional on a ball admits a classical solution. The optimal design in this case is the one with the better conductor placed in a ball around zero, with radius depending on the amount of given material. However, if the domain is changed to a square, this maximization problem does not have a classical solution (Glowinski (1984), Goodman et al. (1986), Lurie & Cherkaev (1984)), there are regions that should be filled with finer and finer mixtures of given materials. Actually, when the domain is a simply connected open set, with a smooth connected boundary, in Casado-Díaz (2015*a*) it was proven that the classical solution for the constant right hand side exists only in a case when the domain is a ball (see also Tartar (1987)). For the minimization of the same functional, the situation is even more complicated, since the classical solution does not exist, not even on a ball (Tartar (1987)). The optimal microstructure in this case with arbitrary right-hand side is explicitly calculated in Casado-Díaz (2015*b*). These results are based on the fact that, for any optimal design problem for the stationary diffusion with one state equation, there exists a relaxed solution which corresponds to a simple laminate in each point of the domain (Murat & Tartar (1985)). As a consequence, the relaxation of the original problem is simplified, written in terms of local proportions of given materials. This can be done for finite number of given materials, even anisotropic ones (Tartar (1995)), although there is no explicit characterization of the G-closure set in this case.

Unfortunately, this approach cannot be generalized to multiple state problems, nor in linearized elasticity problems, because in these problems the higher order sequential laminates appears as solutions (Allaire (2002), Antonić & Vrdoljak (2006), Vrdoljak (2010)). However, recently it was shown that similar results holds in the context of stationary diffusion equation for some classes of multiple state optimal design problems. In the spherically symmetric case Vrdoljak (2016) proved that the solution can be found among simple laminates in the case of maximization of the conic sum of energy functionals on a ball and annulus. Moreover, the solution is classical and in the case of a ball it can be simply calculated using the optimality conditions. In the case of minimizing the same functional, in Burazin (2018) and Burazin & Vrdoljak (2018) the unique solutions are explicitly calculated for various examples on an annulus and a ball. For more complicated domains (or functionals) it is quite unlikely to find an explicit solution (see Goodman et al. (1986)), which imposes a need for various numerical methods.

One of the most popular numerical methods for numerical solution is the optimality criteria method, an iterative method based on necessary conditions for optimality of the relaxed formulation. In the last 15 years, simpler approaches have been used, like the SIMP method (Bendsøe (1995), Bendsøe & Sigmund (2003)), especially in commercial engineering tools. However, the homogenization method has recently experienced a renewed interest because of its suitability for optimizing lattice and porous materials, which have great importance in modern manufacturing techniques like 3-d printing (Allaire et al. (2017*a*,*b*, 2018), Geoffroy-Donders et al. (2018)).

There are numerous results on the optimality criteria method in structural optimization problems (see Allaire (2002), Bendsøe (1995), Kirsch (1981), Rozvany (1989) and references therein). For the single state optimal design problems in the conductivity setting, with the aim of energy optimization, in Murat & Tartar (1985) it was noticed that two different approaches to optimality conditions are needed, one for the minimization and the other one for the maximization of energy. The optimality criteria method for these problems is described in Allaire (2002), where one can also find a convergence proof, based on the proof of Toader (1997). The method is generalized to multiple state optimal design problems in Vrdoljak (2010), and it was noticed that the developed method works properly for the maximization of a conic sum of energies, but fails for the minimization of the same functional. Consequently, there is a need for a new variant of the optimality criteria method suitable for minimizing a conic sum of energies.

Regarding optimal design problems in linearized elasticity, due to the lack of an explicit description of G-closure, even for mixtures of two isotropic phases, it seems that the homogenization method is not an apropriate tool for finding a solution. However, in the case of compliance minimization, the relaxation can be performed on a smaller subset made of sequential laminates, which is explicitly known (Avellaneda (1987), see also Kohn & Lipton (1988) and Milton & Kohn (1988)). Nevertheless, even in the case of the one state equation, the optimal design cannot be found among simple laminates, and the analytic solutions for these problems are not known. On the other hand, for the compliance functional the necessary conditions of optimality are easily derived, which enables the development of an optimality criteria method for a numerical solution. The possible issue is a need for explicit calculation of Hashin-Shtrikman bounds, that naturally arises in the optimality condition in the sense that the optimal design locally (or in a given point) saturates the appropriate Hashin-Shtrikman bound. In the two-dimensional case, bounds on the primal energy are calculated in Allaire & Kohn (1993*a*) for the mixture of two isotropic phases, while in the three-dimensional case it was formally done only for the shape optimization problems, which are considered as above mentioned problems, but with one material being void (Gibianski & Cherkaev (1984, 1987), Allaire & Kohn (1993*b*), Allaire (1994)).

Overview

In Chapter 1 we give a brief overview of homogenization theory. Composite materials are introduced, with emphasis on sequential laminates, a particularly important class of composites. The G-closure set is explicitly characterized in a special case of mixing two isotropic material in a conductivity setting, while for the elasticity setting Hashin-Shtrikman bounds are presented.

A multiple state optimal design problem for the stationary diffusion equation is introduced in Chapter 2. Since it usually does not admit a classical solution, the relaxation by the homogenization method is conducted. Necessary conditions of optimality are derived, which enabled the development of a new variant of the optimality criteria method. Various examples are presented, and based on good behaviour of the algorithm, a convergence of the algorithm is proved in some special cases.

Chapter 3 gives the adaptation of the optimality criteria method developed in the previous chapter to similar problems in linearized elasticity. To succeed in the implementation, an explicit calculation of the lower Hashin-Shtrikman bound on the complementary energy is needed. We give it both in two and three space dimensions and present a number of examples for two-dimensional compliance minimization problems.

Chapter 1

Homogenization theory

The general idea of homogenization is to derive macroscopic (*effective*) properties of microscopically heterogeneous media. As such, it provides a solid ground to the notion of composite materials, fine mixtures of original materials, which are of great importance in optimization theory. There are various theories of homogenization: stohastic theory of homogenization, variational theory of homogenization (Γ-convergence) and H-convergence. The latter was introduced by Spagnolo (1976) under the name of G-convergence, generalized in Tartar (1975), Murat & Tartar (1978), and it is especially well adapted for applications in optimal design problems. Therefore, in this chapter we give some results of H-convergence in the context of linear second order elliptic equations, namely the stationary diffusion equation and linearized elasticity system. More information about homogenization theory and applications in optimal design can be found in Allaire (2002), Murat & Tartar (1985, 1978), and Tartar (2000, 1975).

1.1 H-convergence

We focus on a model problem of thermal (or electrical) conductivity in a medium. Let $\Omega \subseteq \mathbf{R}^d$ be an open bounded set, representing a medium, and $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathrm{M}_d(\mathbf{R}))$ a matrix function containing information about the thermal (or electrical) conductivity of the material that constitutes Ω, called the conductivity matrix. For given external heat (or electric charge) density $f \in \mathrm{H}^{-1}(\Omega)$, the temperature u (or the potential in electrostatics), satisfies the stationary diffusion equaton

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u) = f & \text{in } \Omega \\ u \in \mathrm{H}^1_0(\Omega)\,. \end{cases} \tag{1.1}$$

We assume that $\mathbf{A}$ is measurable, bounded and coercive, i.e. for $0 < \alpha < \beta$, for almost every $\mathbf{x}$, the matrix $\mathbf{A}(\mathbf{x})$ belongs to the set

$$\mathcal{M}_{\alpha,\beta} := \left\{ \mathbf{A} \in \mathrm{M}_d(\mathbf{R}) : (\forall \boldsymbol{\xi} \in \mathbf{R}^d)\; \mathbf{A}\boldsymbol{\xi}\cdot\boldsymbol{\xi} \geq \alpha|\boldsymbol{\xi}|^2,\; \mathbf{A}^{-1}\boldsymbol{\xi}\cdot\boldsymbol{\xi} \geq \frac{1}{\beta}|\boldsymbol{\xi}|^2 \right\}.$$

By application of the Lax-Milgram lemma, the equation (1.1) admits a unique weak solution. Moreover, by invoking the Poincaré inequality:

$$(\exists c > 0)(\forall \varphi \in \mathrm{H}_0^1(\Omega)) \quad \|\varphi\|_{\mathrm{L}^2(\Omega)} \le c\|\nabla\varphi\|_{\mathrm{L}^2(\Omega)},$$

norms $\|\nabla \cdot\|_{\mathrm{L}^2(\Omega)}$ and $\|\cdot\|_{\mathrm{H}_0^1(\Omega)}$ are equivalent. This implies, together with the coercivity of $\mathbf{A}$, the a priori estimate

$$\|\nabla u\|_{\mathrm{L}^2(\Omega)} \le \frac{c}{\alpha}\|f\|_{\mathrm{H}^{-1}}. \tag{1.2}$$

The domain Ω is set to be highly heterogeneous, made up from significantly different materials. Usually, we are not interested in the properties of material in every point of the domain, but rather globally. Mathematically, this could be solved by analyzing a limit (also called *homogenized* or *effective conductivity*) of a sequence of conductivity matrix functions $(\mathbf{A}^n)$ in $\mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$, when n goes to infinity, with the corresponding Dirichlet boundary problems

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^n \nabla u^n) = f \quad \text{in } \Omega \\ u^n \in \mathrm{H}_0^1(\Omega)\,. \end{cases} \tag{1.3}$$

From the a priori estimate (1.2), a sequence of solutions (u^n) of (1.3) is bounded in $\mathrm{H}_0^1(\Omega)$, therefore, there exists a subsequence (denoted the same) such that

$$u^n \rightharpoonup u, \qquad \text{in } \mathrm{H}_0^1(\Omega). \tag{1.4}$$

Furthermore, a sequence of fluxes $\boldsymbol{\sigma}^n = \mathbf{A}^n \nabla u^n$ is bounded in $\mathrm{L}^2(\Omega)$, and there exists a subsequence (denoted the same) such that

$$\boldsymbol{\sigma}^n \rightharpoonup \boldsymbol{\sigma}, \qquad \text{in } \mathrm{L}^2(\Omega; \mathbf{R}^d). \tag{1.5}$$

Since the distributional derivative is continuous in the space of distribution, for the limit $\boldsymbol{\sigma}$ it follows

$$-\mathsf{div}\,\boldsymbol{\sigma} = f. \tag{1.6}$$

One can wonder if there is any relationship between limits $\boldsymbol{\sigma}$ and u, or what kind of equation u satisfies. Perhaps there is also some connection between the sequence $\mathbf{A}^n$ and coefficients in (1.6). What we know for sure is that, since the sequence $(\mathbf{A}^n)$ is bounded in $\mathrm{L}^\infty(\Omega; \mathrm{M}_d(\mathbf{R}))$, there exists a limit $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathrm{M}_d(\mathbf{R}))$ such that

$$\mathbf{A}^n \overset{*}{\rightharpoonup} \mathbf{A}, \qquad \text{in } \mathrm{L}^\infty(\Omega; \mathrm{M}_d(\mathbf{R})) \tag{1.7}$$

(on a subsequence). Can we conclude from (1.4) and (1.7) that $\boldsymbol{\sigma} = \mathbf{A}\nabla u$, i.e. that

$$\mathbf{A}^n \nabla u^n \rightharpoonup \mathbf{A}\nabla u, \qquad \text{in } \mathrm{L}^2(\Omega; \mathbf{R}^d)? \tag{1.8}$$

Example 1.1 In the case when $d = 1$, equation (1.3) reduces to

$$\begin{cases} -\dfrac{d}{dx}\left(a^n \dfrac{d}{dx} u^n\right) = f, \quad \text{in } \langle 0, l\rangle \\ u^n(0) = u^n(l) = 0. \end{cases} \tag{1.9}$$

Since $f \in \mathrm{H}^{-1}(\langle 0, l\rangle)$, there exist $g \in \mathrm{L}^2(\langle 0, l\rangle)$ such that $f = \frac{dg}{dx}$, and therefore

$$\sigma^n = a^n \frac{du^n}{dx} = C^n - g, \quad C^n \in \mathbf{R}.$$

The sequence σ^n is bounded in $\mathrm{L}^2(\langle 0, l\rangle)$, hence C^n is bounded sequence of numbers in $\mathbf{R}$, thus one can extract a subsequence, denoted the same, such that C^n converges to C. Therefore, $\sigma^n \longrightarrow \sigma = C - g$ in $\mathrm{L}^2(\langle 0, l\rangle)$. Moreover, the sequence $\frac{1}{a^n}$ is bounded in $\mathrm{L}^\infty(\langle 0, l\rangle)$, which implies $\frac{1}{a^n} \overset{*}{\rightharpoonup} \frac{1}{a}$ in $\mathrm{L}^\infty(\langle 0, l\rangle)$. Since the dual product of a weakly and strongly converging sequence converges to the dual product of its limits, from (1.4) it follows

$$\frac{1}{a}\sigma = \frac{du}{dx} \tag{1.10}$$

and, as a consequence, function u is a solution of equation

$$\begin{cases} -\dfrac{d}{dx}\left(a \dfrac{d}{dx} u\right) = f, \quad \text{in } \langle 0, l\rangle \\ u(0) = u(l) = 0. \end{cases} \tag{1.11}$$

In general, (1.8) is not true, since the dual product of two weakly converging sequences does not necessarily converge to the dual product of the weak limits. Homogenization theory studies an appropriate topology on the space of coefficients for which a limit function u is also a solution of an equation of the same type.

Definition 1.1 A sequence of matrices $\mathbf{A}^n \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$ is said to H-converge to a homogenized limit, or H-limit, $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha',\beta'})$ if, for any right hand side $f \in \mathrm{H}^{-1}(\Omega)$, the sequence u^n of solutions of

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^n \nabla u^n) = f \quad \text{in } \Omega \\ u^n \in \mathrm{H}^1_0(\Omega) \end{cases}$$

satisfies

$$\begin{cases} u^n \rightharpoonup u \text{ in } \mathrm{H}^1_0(\Omega) \\ \mathbf{A}^n \nabla u^n \rightharpoonup \mathbf{A}\nabla u \text{ in } \mathrm{L}^2(\Omega; \mathbf{R}^d)\,, \end{cases}$$

and, consequently, u is the solution of the homogenized equation

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u) = f \quad \text{in } \Omega \\ u \in \mathrm{H}^1_0(\Omega)\,. \end{cases}$$

The above definition is justified by the following compactness result (Murat & Tartar (1978)), where the existence of an H-limit is proved.

Theorem 1.1 For any sequence $\mathbf{A}^n \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$ there exists a subsequence which H-converges to some $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$.

□

H-convergence has some nice properties, given in the following remark. For proofs and other properties we refer to Tartar (2000, 1995), and Allaire (2002).

Remark 1.1.

(i) H-convergence is local, meaning that the value of the homogenized matrix $\mathbf{A}$ in some open set does not depend on the values of the sequence $(\mathbf{A}^n)$ outside this set. This also implies that the H-limit is unique.

(ii) H-convergence is indifferent with the respect to the choice of boundary conditions (for simplicity, we took Dirichlet boundary conditions).

(iii) The topology induced by the H-convergence (H-topology) is metrizable.

(iv) If a sequence $\mathbf{A}^n \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$ either converges strongly to a limit $\mathbf{A}$ in $\mathrm{L}^1(\Omega; \mathrm{Sym_d})$, or converges to $\mathbf{A}$ almost everywhere in Ω, then $\mathbf{A}^n$ H-converges to $\mathbf{A}$.

(v) The homogenized limit $\mathbf{A}$ of a sequence of symmetric matrices $\mathbf{A}^n \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$ is bounded below by the harmonic mean, and bounded above by the arithmetic mean, i.e.

$$\underline{\mathbf{A}}\boldsymbol{\xi} \cdot \boldsymbol{\xi} \le \mathbf{A}\boldsymbol{\xi} \cdot \boldsymbol{\xi} \le \overline{\mathbf{A}}\boldsymbol{\xi} \cdot \boldsymbol{\xi}, \quad \boldsymbol{\xi} \in \mathbf{R}^d,$$

where $\underline{\mathbf{A}}$ and $\overline{\mathbf{A}}^{-1}$ are weak-$*$ limits of $\mathbf{A}^n$ and $(\mathbf{A}^n)^{-1}$ in $\mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$, respectively. Moreover, $\underline{\mathbf{A}} = \overline{\mathbf{A}}$ if and only if $\mathbf{A}^n$ converges almost everywhere in Ω, or converges strongly in $\mathrm{L}^p(\Omega)$ for some p, $1 \le p \le \infty$.

Note that in general, there is no explicit formula for the H-limit. However, there are some special cases where the formula can be obtained, such as periodic or laminated structures. In the periodic case, for a unit cell $Y = [0, 1]^d$, we define the Lebesgue and Sobolev space of Y-periodic function in $\mathbf{R}^d$ by

$$\mathrm{L}^p_\#(Y) := \{f \in \mathrm{L}^p_{\mathrm{loc}}(\mathbf{R}^d) : f \text{ is } Y\text{-periodic}\},$$

$1 \le p \le \infty$, and

$$\mathrm{H}^1_\#(Y) := \{f \in \mathrm{H}^1_{\mathrm{loc}}(\mathbf{R}^d) : f \text{ is } Y\text{-periodic}\}.$$

Both $(\mathrm{L}^p_\#(Y), \|\cdot\|_{\mathrm{L}^p(Y)})$ and $(\mathrm{H}^1_\#(Y), \|\cdot\|_{\mathrm{H}^1(Y)})$ are Banach spaces. We also define the quotient space $\mathrm{H}^1_\#(Y)/\mathbf{R}$ as the space of classes of functions in $\mathrm{H}^1_\#(Y)$ equal up to an additive constant. These definitions easily extend to vector or matrix valued functions. If we take a Y-periodic matrix function $\mathbf{A}_0 \in \mathrm{L}^\infty_\#(Y; \mathrm{M}_d(\mathbf{R}))$ such that $\mathbf{A}_0(\mathbf{x})$ belongs to $\mathcal{M}_{\alpha,\beta}$ for a.e. $\mathbf{x} \in Y$, then H-limit of the sequence

$$\mathbf{A}^n(\mathbf{x}) = \mathbf{A}_0(n\mathbf{x}), \quad \mathbf{x} \in \Omega \tag{1.12}$$

is given in the following theorem (Cioranescu & Donato (1999), Allaire (2002)).

Theorem 1.2 The sequence $\mathbf{A}^n$ defined by (1.12) H-converges to a constant homogenized matrix $\mathbf{A} \in \mathcal{M}_{\alpha,\beta}$, defined by its entries

$$\mathbf{A}_{ij} = \int_Y \mathbf{A}(\mathbf{y})(\mathsf{e}_i + \nabla w_i(\mathbf{y})) \cdot (\mathsf{e}_j + \nabla w_j(\mathbf{y}))\, d\mathbf{y},$$

where e_i, $i = 1, \dots, d$ is the canonical basis of $\mathbf{R}^d$, and w_i, $i = 1, \dots, d$ is the family of unique solutions in $\mathrm{H}^1_\#(Y)/\mathbf{R}$ of the cell problems

$$\begin{cases} -\mathsf{div}\, \mathbf{A}(\mathbf{y})(\mathsf{e}_i + \nabla w_i(\mathbf{y})) = 0 & \text{in } Y \\ w_i \text{ is } Y\text{- periodic}, & i = 1, \dots, d. \end{cases}$$

□

Actually, it can be proven that at each point, a general H-limit can be attained as the limit of a sequence of periodic homogenized matrices, stated in theorem below (see, e.g., Zhikov et al. (1994), Allaire (2002)).

Theorem 1.3 Let $\mathbf{A}^n$ be a sequence of matrices in $\mathrm{L}^\infty_\#(Y; \mathcal{M}_{\alpha,\beta})$ which H-converges to a limit $\mathbf{A}$. For any $\mathbf{x} \in \Omega$ and any sufficiently large $m \in \mathbf{N}$, let $\mathbf{A}_{\mathbf{x},n,m}$ be the periodic homogenized matrix defined by its entries

$$(\mathbf{A}_{\mathbf{x},n,m})_{ij} = \int_Y \mathbf{A}^n \left(\mathbf{x} + \frac{\mathbf{y}}{m}\right) (\mathsf{e}_i + \nabla w^i_{\mathbf{x},n,m}) \cdot (\mathsf{e}_i + \nabla w^i_{\mathbf{x},n,m})\, d\mathbf{y},$$

where $w^i_{\mathbf{x},n,m}$, $i = 1, \dots, d$ is the family of unique solutions in $\mathrm{H}_\#(Y)/\mathbf{R}$ of the cell problems

$$\begin{cases} -\mathsf{div}\, \mathbf{A}^n \left(\mathbf{x} + \frac{\mathbf{y}}{m}\right) (\mathsf{e}_i + \nabla w^i_{\mathbf{x},n,m}(\mathbf{y})) = 0 & \text{in } Y \\ w^i_{\mathbf{x},n,m} \text{ is } Y\text{- periodic}, & i = 1, \dots, d. \end{cases}$$

There exists a diagonal subsequence m' of the sequence m such that

$$\mathbf{A}_{\mathbf{x},n(m'),m'} \longrightarrow \mathbf{A}(\mathbf{x}), \quad m' \longrightarrow \infty,$$

for almost every $\mathbf{x} \in \Omega$.

□

Theorem 1.3 implies that the set of all periodic H-limits is dense in the set of pointwise values of general H-limits. In particular, for $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{M}_{\alpha,\beta}$ and $\chi^n \in \mathrm{L}^\infty(\Omega; \{0,\, 1\})$, consider a sequence

$$\mathbf{A}^n = \chi^n \mathbf{A}_1 + (1 - \chi^n)\mathbf{A}_2$$

which H-converges to a limit $\mathbf{A}$ in $\mathrm{L}^\infty(\Omega, \mathcal{M}_{\alpha,\beta})$. Since χ^n is bounded in $\mathrm{L}^\infty(\Omega; \{0,\, 1\})$, there exists $\theta \in \mathrm{L}^\infty(\Omega; [0,1])$ such that, up to a subsequence, χ^n weak-$*$ converges to θ in $\mathrm{L}^\infty(\Omega; [0,1])$. This limit θ is the local proportion of material $\mathbf{A}_1$ in the H-limit $\mathbf{A}$. On the other hand, for the periodic homogenized limit $\mathbf{A}_{\mathbf{x},n,m}$ defined in Theorem 1.3, which approximates $\mathbf{A}$, the local proportion of material $\mathbf{A}_1$ is given by $\theta_{\mathbf{x},n,m} = \int_Y \chi^n \left(\mathbf{x} + \frac{\mathbf{y}}{m}\right) d\mathbf{y}$. It can also be proved that there exists a diagonal subsequence m' of the sequence m such that

$$\theta_{\mathbf{x},n(m'),m'} \longrightarrow \theta(\mathbf{x}), \quad m' \longrightarrow \infty,$$

for almost every $\mathbf{x} \in \Omega$, which implies that in this approximation of the H-limit $\mathbf{A}$ by a sequence of matrices obtained by periodic homogenization, the local proportions of materials $\mathbf{A}_1$ and $\mathbf{A}_2$ are also preserved in the limit.

1.2 Composite materials

In this section we shall give the main features of composite materials: heterogeneous materials obtained by mixing several phases on a very fine scale. Homogenization theory gives an appropriate tool for the definition of composite material as an H-limit of a sequence of increasingly finer mixtures of the constituent materials. This also enables us to determine their effective properties. We shall focus on mixtures obtained by mixing two materials $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{M}_{\alpha,\beta}$, and since the conductivity matrix is always symmetric, we assume $\mathbf{A}_1, \mathbf{A}_2 \in \mathrm{Sym}_d$, where Sym_d denotes the space of all symmetric $d \times d$ matrices. Moreover, we assume a perfect bonding between the two materials, meaning that temperature and heat flux are continuous at the interface. We start with the definition of a composite (or generalized) material.

Definition 1.2 Let $\chi^n \in \mathrm{L}^\infty(\Omega; \{0,1\})$ be a sequence of characteristic functions and $\mathbf{A}^n$ sequence of tensors defined by

$$\mathbf{A}^n(\mathbf{x}) = \chi^n(\mathbf{x})\mathbf{A}_1 + (1 - \chi^n(\mathbf{x}))\mathbf{A}_2, \quad \mathbf{x} \in \Omega. \tag{1.13}$$

Assume that there exist $\theta \in \mathrm{L}^\infty(\Omega; [0,1])$ and $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$ such that

$$\chi^n \overset{*}{\rightharpoonup} \theta \quad \text{ in } \mathrm{L}^\infty(\Omega; [0,1])$$

and

$$\mathbf{A}^n \overset{H}{\rightharpoonup} \mathbf{A} \quad \text{ in } \mathrm{L}^\infty(\Omega, \mathcal{M}_{\alpha,\beta})$$

in the sense of Definition 1.1. Then the H-limit $\mathbf{A}$ is said to be the homogenized tensor of a two-phase composite material obtained by mixing materials $\mathbf{A}_1$ and $\mathbf{A}_2$ in proportions θ and $(1-\theta)$, respectively, with a microstructure defined by the sequence (χ^n).

Recalling Theorem 1.1, the H-limit of (1.13) exists, at least on a subsequence, and, therefore, the above definition is justified. As it can be seen from the definition, a composite material is defined by phases $\mathbf{A}_1$ and $\mathbf{A}_2$, and the microstructure (χ^n). Let us, for a given density function $\theta \in \mathrm{L}^\infty(\Omega;[0,1])$, define a set of all possible homogenized conductivities which can be obtained by mixing phases $\mathbf{A}_1$ and $\mathbf{A}_2$ in proportions θ and $(1-\theta)$, respectively,

$$\begin{aligned} \mathcal{G}_\theta := \{\mathbf{A} \in \mathrm{L}^\infty(\Omega;\mathcal{M}_{\alpha,\beta}) :\ & \text{there exists } (\chi^n) \text{ in } \mathrm{L}^\infty(\Omega;\{0,1\}) \text{ such that} \\ \chi^n \xrightarrow{*} \theta \ \text{ in } \mathrm{L}^\infty(\Omega;[0,1]) & \text{ and } \chi^n\mathbf{A}_1 + (1-\chi^n)\mathbf{A}_2 \xrightarrow{H} \mathbf{A}\}. \end{aligned} \tag{1.14}$$

Lemma 1.4 The set $\mathcal{G}_\theta$ is a closed set in $\mathrm{L}^\infty(\Omega,\mathcal{M}_{\alpha,\beta})$ with respect to H-topology.

Proof. Since H-topology is metrizable by Remark 1.1(iii), we can use Cantor diagonal process. Let $\mathbf{A}^n \in \mathcal{G}_\theta$ such that $\mathbf{A}^n \xrightarrow{H} \mathbf{A}$ and let us prove that $\mathbf{A} \in \mathcal{G}_\theta$. From the definition of the set $\mathcal{G}_\theta$, matrices $\mathbf{A}^n$ can also be understood as H-limits of some sequences $\mathbf{A}^n_k$, more precisely there exist sequences $\chi^n_k \in \mathrm{L}^\infty(\Omega;\{0,1\})$ and $\mathbf{A}^n_k = \chi^n_k\mathbf{A}_1 + (1-\chi^n_k)\mathbf{A}_2$ such that

$$\begin{aligned} \chi^n_k &\xrightarrow{*} \theta \quad \text{in } \mathrm{L}^\infty(\Omega;[0,1]), \\ \mathbf{A}^n_k &\xrightarrow{H} \mathbf{A}^n \quad \text{in } \mathrm{L}^\infty(\Omega;\mathcal{M}_{\alpha,\beta}). \end{aligned}$$

Then, Cantor's diagonalization process gives sequences $(\chi^k) \in \mathrm{L}^\infty(\Omega;\{0,1\})$ and $\mathbf{A}^k = \chi^k\mathbf{A}_1 + (1-\chi^k)\mathbf{A}_2$ such that $\chi^k \xrightarrow{*} \theta$ in $\mathrm{L}^\infty(\Omega;[0,1])$ and $\mathbf{A}^k \xrightarrow{H} \mathbf{A}$ in $\mathrm{L}^\infty(\Omega,\mathcal{M}_{\alpha,\beta})$, which implies $\mathbf{A} \in \mathcal{G}_\theta$. □

The set $\mathcal{G}_\theta$ is also known as *G-closure set*, referring to the problem of finding the closure of a set containing matrix functions of form $\chi^n(\cdot)\mathbf{A}_1 + (1-\chi^n(\cdot))\mathbf{A}_2$, for χ^n being the weak-$*$ convergent sequence of characteristic functions, under G- or H-convergence (*G-closure problem*). We proceed with the study of this set. For a constant $\theta \in [0,1]$, we denote by P_θ all homogenized tensors obtained by periodic homogenization of materials $\mathbf{A}_1$ and $\mathbf{A}_2$, with proportion θ of material $\mathbf{A}_1$. More precisely, an element $\mathbf{A} \in P_\theta$ is a constant H-limit of $\mathbf{A}^n = \chi^n\mathbf{A}_1 + (1-\chi^n)\mathbf{A}_2$, where $\chi^n(\mathbf{x}) = \chi(n\mathbf{x})$, for $\chi \in \mathrm{L}^\infty(\Omega;[0,1])$, and $\int_Y \chi(\mathbf{y})\,d\mathbf{y} = \theta$. From Remark 1.1(v), the set P_θ is bounded, but not necessarily closed in Sym_d. Therefore, let us denote by G_θ its closure in Sym_d, i.e.

$$G_\theta = \mathrm{Cl}\, P_\theta. \tag{1.15}$$

Remark 1.2. Notice that the set G_θ is set of all possible two-phase composite materials at fixed volume fraction θ. Indeed, for $\theta \in [0,1]$, if a matrix $\mathbf{A}$ is an H-limit of sequence $\chi^n\mathbf{A}_1+(1-\chi^n)\mathbf{A}_2$, where $\chi^n \xrightarrow{*} \theta$ in $\mathrm{L}^\infty(\Omega;[0,1])$, then Theorem 1.3 implies that $\mathbf{A} \in G_\theta$. Vice versa, if $\mathbf{A} \in G_\theta$, then there exist a sequence $\mathbf{A}^n \in P_\theta$ that converges strongly to $\mathbf{A}$. Moreover, $\mathbf{A}^n$ are H-limits obtained by periodic homogenization of materials $\mathbf{A}_1$ and $\mathbf{A}_2$ with proportion θ of material $\mathbf{A}_1$. By extracting a diagonal subsequence the assertion is proven.

Theorem 1.5 For any function $\theta \in \mathrm{L}^\infty(\Omega;[0,1])$, the G-closure set is characterized by

$$\mathcal{G}_\theta = \left\{\mathbf{A} \in \mathrm{L}^\infty(\Omega;\mathrm{Sym}_d) : \mathbf{A}(\mathbf{x}) \in G_{\theta(\mathbf{x})},\ \text{a. e. } \mathbf{x} \in \Omega\right\}, \tag{1.16}$$

where for any real number $\theta \in [0,1]$, the set G_θ is defined by (1.15).

Before providing the proof of the above theorem, note that it in fact asserts that there are no global, but only local properties of homogenized tensors in $\mathcal{G}_\theta$. Therefore, when studying composite materials, it is enough to characterize set G_θ, and furthermore, we can assume that the effective tensors are constant and obtained by periodic homogenization. We shall also refer to G_θ as the G-closure set in the sequel. For the proof of the above theorem, we shall need the following lemma (Allaire 2002, Lemma 2.1.7).

Lemma 1.6 There exist positive constants $c > 0$ and $\delta > 0$ such that, for any $\theta_1\,\theta_2 \in [0,1]$,

$$d(G_{\theta_1}, G_{\theta_2}) \le c|\theta_1 - \theta_2|^\delta,$$

where d denotes the Hausdorff distance for compact subsets of Sym_d, defined by

$$d(K_1,K_2) = \max\left\{\max_{x_1\in K_1}\min_{x_2\in K_2} d(x_1,x_2), \max_{x_2\in K_2}\min_{x_1\in K_1} d(x_2,x_1)\right\}, \quad K_1,K_2 \subseteq \mathrm{Sym}_d.$$

□

Proof of Theorem 1.5. We denote

$$\mathcal{A}_\theta = \{\mathbf{A} \in \mathrm{L}^\infty(\Omega;\mathrm{Sym}_d) : \mathbf{A}(\mathbf{x}) \in G_{\theta(\mathbf{x})},\ \text{a.e. } \mathbf{x} \in \Omega\}$$

and prove $\mathcal{G}_\theta = \mathcal{A}_\theta$. First, let us assume that $\mathbf{A} \in \mathcal{G}_\theta$. From Theorem 1.3 (and comments below it), there exists a sequence $\theta^n \in \mathrm{L}^\infty(\Omega;[0,1])$ and $\mathbf{A}^n \in \mathrm{L}^\infty(\Omega;\mathcal{M}_{\alpha,\beta})$ such that

$$\theta^n \longrightarrow \theta, \quad \mathbf{A}^n \longrightarrow \mathbf{A}, \text{ a.e. on } \Omega$$

and $\mathbf{A}^n(\mathbf{x}) \in P_{\theta^n(\mathbf{x})} \subseteq G_{\theta^n(\mathbf{x})}$. Since $G_{\theta(\mathbf{x})}$ is compact, we can compute a distance from $\mathbf{A}(\mathbf{x})$ to $G_{\theta(\mathbf{x})}$ and it follows

$$\begin{aligned} d(\mathbf{A}(\mathbf{x}), G_{\theta(\mathbf{x})}) &\le d(\mathbf{A}(\mathbf{x}), \mathbf{A}^n(\mathbf{x})) + d(\mathbf{A}^n(\mathbf{x}), G_{\theta(\mathbf{x})}) \\ &\le d(\mathbf{A}(\mathbf{x}), \mathbf{A}^n(\mathbf{x})) + c|\theta^n(\mathbf{x}) - \theta(\mathbf{x})|^\delta, \end{aligned}$$

where the last inequality follows from Lemma 1.6. Passing to the limit when $n \longrightarrow \infty$, we get $\mathbf{A}(\mathbf{x}) \in G_{\theta(\mathbf{x})}$ a.e. $\mathbf{x} \in \Omega$, i.e. $\mathcal{G}_\theta \subseteq \mathcal{A}_\theta$.

Vice versa, if $\mathbf{A} \in \mathcal{A}_\theta$, we shall prove that $\mathbf{A} \in \mathcal{G}_\theta$ by approximating it by a piecewise constant functions. For $n \in \mathbf{N}$, let $(\omega_j^n)_{1\le j\le n}$ be a family of disjoint open sets covering Ω such that $\max_{1\le j\le n} \operatorname{diam} \omega_j^n \xrightarrow{n\to\infty} 0$. Define

$$\theta_j^n := \frac{1}{|\omega_j^n|}\int_{\omega_j^n} \theta(\mathbf{x})\,d\mathbf{x} \text{ and } \theta^n(\mathbf{x}) := \sum_{j=1}^n \theta_j^n \chi_j^n(\mathbf{x}),$$

where χ_j^n is the characteristic function of ω_j^n. Then

$$\theta^n \longrightarrow \theta, \text{ in } \mathrm{L}^p(\Omega;[0,1]),\ 1 \le p < \infty. \tag{1.17}$$

Moreover, we construct a sequence $\tilde{\mathbf{A}}^n \in \mathrm{L}^\infty(\Omega;\mathrm{Sym}_d)$ on ω_j^n as a projection of $\mathbf{A}$ on $G_{\theta_j^n}$. Lemma 1.6 implies

$$|\mathbf{A}(\mathbf{x}) - \tilde{\mathbf{A}}^n(\mathbf{x})| \le d(G_{\theta(\mathbf{x})}, G_{\theta_j^n}) \le c|\theta(\mathbf{x}) - \theta_j^n(\mathbf{x})|^\delta, \text{ a.e. } \mathbf{x} \in \omega_j^n.$$

Therefore, using (1.17), $\tilde{\mathbf{A}}^n$ converges strongly to $\mathbf{A}$ in $\mathrm{L}^p(\Omega;\mathrm{Sym}_d)$, $1 \le p < \infty$. However, each $\tilde{\mathbf{A}}^n$ is not piecewise constant. Thus we define

$$\hat{\mathbf{A}}_j^n := \frac{1}{|\omega_j^n|}\int_{\omega_j^n} \tilde{\mathbf{A}}^n\,d\mathbf{x} \text{ and } \hat{\mathbf{A}}^n(\mathbf{x}) := \sum_{j=1}^n \hat{\mathbf{A}}_j^n \chi_j^n(\mathbf{x}).$$

It can easily be seen that $\hat{\mathbf{A}}^n$ converges strongly to $\mathbf{A}$ in $\mathrm{L}^p(\Omega;\mathrm{Sym}_d)$, $1 \le p < \infty$. However, there is no guarantee that each $\hat{\mathbf{A}}_j^n$ is an H-limit, i.e. that it belongs to $G_{\theta_j^n}$. Therefore, we define $\mathbf{A}_j^n$ as the projection of $\hat{\mathbf{A}}_j^n$ on $G_{\theta_j^n}$ and

$$\mathbf{A}^n(\mathbf{x}) := \sum_{j=1}^n \mathbf{A}_j^n \chi_j^n(\mathbf{x}).$$

By definition $\mathbf{A}^n \in \mathcal{G}_{\theta^n}$, and it remains to prove strong convergence of $\mathbf{A}^n$ to $\mathbf{A}$ in $\mathrm{L}^1(\Omega;\mathrm{Sym}_d)$. For $\mathbf{x} \in \omega_j^n$,

$$|\mathbf{A}^n(\mathbf{x}) - \hat{\mathbf{A}}^n(\mathbf{x})| = |\mathbf{A}_j^n - \hat{\mathbf{A}}_j^n| \le |\tilde{\mathbf{A}}^n(\mathbf{x}) - \hat{\mathbf{A}}^n(\mathbf{x})|.$$

Therefore,

$$\begin{aligned} |\mathbf{A}^n(\mathbf{x}) - \mathbf{A}(\mathbf{x})| &\le |\mathbf{A}^n(\mathbf{x}) - \hat{\mathbf{A}}^n(\mathbf{x})| + |\hat{\mathbf{A}}^n(\mathbf{x}) - \tilde{\mathbf{A}}^n(\mathbf{x})| + |\tilde{\mathbf{A}}^n(\mathbf{x}) - \mathbf{A}(\mathbf{x})| \\ &\le 2|\hat{\mathbf{A}}^n(\mathbf{x}) - \tilde{\mathbf{A}}^n(\mathbf{x})| + |\tilde{\mathbf{A}}^n(\mathbf{x}) - \mathbf{A}(\mathbf{x})|, \quad \mathbf{x} \in \omega_j^n. \end{aligned}$$

Since $\tilde{\mathbf{A}}^n$ converges strongly to $\mathbf{A}$ in $\mathrm{L}^p(\Omega;\mathrm{Sym}_d)$, and the difference $\hat{\mathbf{A}}^n - \tilde{\mathbf{A}}^n$ converges strongly to zero in the same space, it follows

$$\mathbf{A}^n \longrightarrow \mathbf{A} \text{ in } \mathrm{L}^p(\Omega; \mathrm{Sym}_d),$$

for any $1 \leq p < \infty$. Strong convergence implies H-convergence (see Remark 1.1(iv)), and since every $\mathbf{A}^n$ belongs to $\mathcal{G}_{\theta^n}$, extracting a diagonal subsequence proves that $\mathbf{A}$ is an H-limit. Furthermore, the strong convergence of θ^n to θ implies that $\mathbf{A} \in \mathcal{G}_\theta$, as desired. □

We can conclude that in order to describe the G-closure set, it is sufficient to know the set G_θ, for given $\theta \in [0, 1]$. As we shall see later, the set G_θ is explicitly described in the case of mixing two isotropic materials. In general, there exist optimal bounds on the set G_θ, while its explicit characterization is still an open problem.

1.2.1 Laminated composites and Hashin-Shtrikman bounds

An important subset of composite materials are laminated composite materials. They have an important role in this dissertation, hence we give a brief overview of laminated composites in this subsection. Let us remark that, in the sequel, we shall abuse the notation $f(\mathbf{x})$ for both the mapping $\mathbf{x} \mapsto f(\mathbf{x})$ and the value of the mapping in a point $\mathbf{x}$. We begin with *simple laminates*, layered mixtures of two materials with prescribed proportion of the first material and layers perpendicular to the lamination direction. Let $\mathsf{e} \in \mathbf{R}^d$ be a unit vector, and $\theta(\mathbf{x} \cdot \mathsf{e})$ the weak-$*$ limit of the sequence $\chi^n(\mathbf{x} \cdot \mathsf{e})$ in $\mathrm{L}^\infty(\Omega; [0, 1])$. Then the sequence $\mathbf{A}^n(\mathbf{x} \cdot \mathsf{e}) = \chi^n(\mathbf{x} \cdot \mathsf{e})\mathbf{A}_1 + (1 - \chi^n(\mathbf{x} \cdot \mathsf{e}))\mathbf{A}_2$ H-converges to $\mathbf{A}(\mathbf{x} \cdot \mathsf{e})$, where

$$\mathbf{A} = \theta\mathbf{A}_1 + (1-\theta)\mathbf{A}_2 - \frac{\theta(1-\theta)}{(1-\theta)\mathbf{A}_1\mathsf{e}\cdot\mathsf{e} + \theta\mathbf{A}_2\mathsf{e}\cdot\mathsf{e}}(\mathbf{A}_2 - \mathbf{A}_1)(\mathsf{e}\otimes\mathsf{e})(\mathbf{A}_2 - \mathbf{A}_1), \quad (1.18)$$

where symbol $\otimes$ indicates the tensor product of two vectors (Murat & Tartar (1978)). We say that $\mathbf{A}$ is a simple laminate in the lamination direction e.

Example 1.2 Let us consider a simple laminate made of two isotropic materials $\mathbf{A}_1 = \alpha\mathbf{I}$ and $\mathbf{A}_2 = \beta\mathbf{I}$, for $0 < \alpha < \beta$. For a given proportion θ of the first material, and the lamination direction e formula (1.18) simplifies to

$$\mathbf{A} = \lambda_\theta^+\mathbf{I} - (\lambda_\theta^+ - \lambda_\theta^-)\mathsf{e}\otimes\mathsf{e},$$

where $\lambda_\theta^+ = \theta\alpha + (1-\theta)\beta$ and $\lambda_\theta^- = \left(\frac{\theta}{\alpha} + \frac{1-\theta}{\beta}\right)^{-1}$. This simple laminate has one eigenvalue equal to λ_θ^- in the lamination direction, while all others are equal to λ_θ^+ (which implies minimal conductivity of a homogenized material in the lamination direction).

We can vary the proportion θ and vector e in order to get a whole family of simple laminates. Furthermore, we can also laminate those laminates. More precisely, for a unit vector $\mathsf{f} \in \mathbf{R}^d$, two simple laminates $\mathbf{A}$ and $\mathbf{B}$ and their proportions ρ and $(1-\rho)$, formula (1.18) yields a new composite $\mathbf{C}$, given by

$$\mathbf{C} = \rho\mathbf{A} + (1-\rho)\mathbf{B} - \frac{\rho(1-\rho)}{(1-\rho)\mathbf{A}\mathsf{f}\cdot\mathsf{f} + \rho\mathbf{B}\mathsf{f}\cdot\mathsf{f}}(\mathbf{B}-\mathbf{A})(\mathsf{f}\otimes\mathsf{f})(\mathbf{B}-\mathbf{A}). \tag{1.19}$$

However, in many applications, the so-called *sequential laminates* seem essential, where the lamination process is iterated using the same pure phase in each lamination. In order to define sequential laminates, we first rewrite the lamination formula (1.18) in a more convenient way.

Lemma 1.7 When the matrix $(\mathbf{A}_1 - \mathbf{A}_2)$ is invertibile, formula (1.18) is equivalent to

$$\theta(\mathbf{A}-\mathbf{A}_2)^{-1} = (\mathbf{A}_1-\mathbf{A}_2)^{-1} + \frac{1-\theta}{\mathbf{A}_2\mathsf{e}\cdot\mathsf{e}}\mathsf{e}\otimes\mathsf{e}. \tag{1.20}$$

Proof. Using the tensor product property $\mathbf{M}(\mathsf{e}\otimes\mathsf{e})\mathbf{M} = (\mathbf{M}\mathsf{e})\otimes(\mathbf{M}^\top\mathsf{e})$, for $\mathbf{M}\in \mathrm{M}_d(\mathbf{R})$, $\mathsf{e}\in\mathbf{R}^d$, we rewrite the formula (1.18) as

$$\theta(\mathbf{A}-\mathbf{A}_2)^{-1} = \left(\mathbf{A}_1 - \mathbf{A}_2 - \frac{1-\theta}{(1-\theta)\mathbf{A}_1\mathsf{e}\cdot\mathsf{e} + \theta\mathbf{A}_2\mathsf{e}\cdot\mathsf{e}}((\mathbf{A}_2-\mathbf{A}_1)\mathsf{e})\otimes\Big((\mathbf{A}_2-\mathbf{A}_1)^\top\mathsf{e}\Big)\right)^{-1}.$$

Moreover, if a matrix $\mathbf{M}$ is regular, and $1 + c(\mathbf{M}\mathsf{e}\cdot\mathsf{e}) \neq 0$, $c\in\mathbf{R}$, then the matrix $\mathbf{M} + c(\mathbf{M}\mathsf{e})\otimes(\mathbf{M}^\top\mathsf{e})$ is regular and it can be easily shown that

$$\Big(\mathbf{M} + c(\mathbf{M}\mathsf{e})\otimes(\mathbf{M}^\top\mathsf{e})\Big)^{-1} = \mathbf{M}^{-1} - \frac{c}{1+c(\mathbf{M}\mathsf{e}\cdot\mathsf{e})}\mathsf{e}\otimes\mathsf{e}.$$

Using the above with $\mathbf{M} = \mathbf{A}_1 - \mathbf{A}_2$, and $c = -\dfrac{1-\theta}{(1-\theta)\mathbf{A}_1\mathsf{e}\cdot\mathsf{e} + \theta\mathbf{A}_2\mathsf{e}\cdot\mathsf{e}}$, a simple computation gives the formula (1.20). $\square$

The advantage of formula (1.20) is that it can be easily iterated. More precisely, let $\mathbf{B}_1$ be a composite obtained by a single lamination of materials $\mathbf{A}_1$ and $\mathbf{A}_2$ in proportions θ_1 and $(1-\theta_1)$ respectively, with layers orthogonal to a unit vector $\mathsf{e}_1\in\mathbf{R}^d$, with representation

$$\theta_1(\mathbf{B}_1-\mathbf{A}_2)^{-1} = (\mathbf{A}_1-\mathbf{A}_2)^{-1} + \frac{1-\theta_1}{\mathbf{A}_2\mathsf{e}_1\cdot\mathsf{e}_1}\mathsf{e}_1\otimes\mathsf{e}_1. \tag{1.21}$$

This simple laminate $\mathbf{B}_1$ can again be laminated with the phase $\mathbf{A}_2$, in proportions θ_2 and $(1-\theta_2)$ and lamination direction e_2. The homogenized tensor $\mathbf{B}_2$ is given by

$$\begin{aligned}\theta_2(\mathbf{B}_2-\mathbf{A}_2)^{-1} &= (\mathbf{B}_1-\mathbf{A}_2)^{-1} + \frac{1-\theta_2}{\mathbf{A}_2\mathsf{e}_2\cdot\mathsf{e}_2}\mathsf{e}_2\otimes\mathsf{e}_2\\ &= \frac{1}{\theta_1}\left((\mathbf{A}_1-\mathbf{A}_2)^{-1} + \frac{1-\theta_1}{\mathbf{A}_2\mathsf{e}_1\cdot\mathsf{e}_1}\mathsf{e}_1\otimes\mathsf{e}_1\right) + \frac{1-\theta_2}{\mathbf{A}_2\mathsf{e}_2\cdot\mathsf{e}_2}\mathsf{e}_2\otimes\mathsf{e}_2,\end{aligned}$$

that is

$$\theta_1\theta_2(\mathbf{B}_2 - \mathbf{A}_2)^{-1} = (\mathbf{A}_1 - \mathbf{A}_2)^{-1} + \frac{1-\theta_1}{\mathbf{A}_2\mathsf{e}_1 \cdot \mathsf{e}_1}\mathsf{e}_1 \otimes \mathsf{e}_1 + \frac{\theta_1(1-\theta_2)}{\mathbf{A}_2\mathsf{e}_2 \cdot \mathsf{e}_2}\mathsf{e}_2 \otimes \mathsf{e}_2.$$

Material $\mathbf{B}_2$ is called a rank-2 sequential laminate with matrix material $\mathbf{A}_2$. We can continue this lamination process with the same phase $\mathbf{A}_2$, and in the end obtain a rank-p sequential laminate $\mathbf{B}_p$ by laminating $\mathbf{B}_{p-1}$ with $\mathbf{A}_2$ in proportions θ_p and $(1-\theta_p)$, respectively, in direction e_p. The homogenized tensor $\mathbf{B}_p$ is defined by

$$\theta_p(\mathbf{B}_p - \mathbf{A}_2)^{-1} = (\mathbf{B}_{p-1} - \mathbf{A}_2)^{-1} + \frac{1-\theta_p}{\mathbf{A}_2\mathsf{e}_p \cdot \mathsf{e}_p}\mathsf{e}_p \otimes \mathsf{e}_p.$$

By proceeding analogously as for rank-2 laminate, i.e. by entering the formulas for $\mathbf{B}_{p-1}, \ldots, \mathbf{B}_1$, the above formula becomes

$$\left(\prod_{i=1}^{p}\theta_i\right)(\mathbf{B}_p - \mathbf{A}_2)^{-1} = (\mathbf{A}_1 - \mathbf{A}_2)^{-1} + \sum_{i=1}^{p}\left((1-\theta_i)\prod_{j=1}^{i-1}\theta_j\right)\frac{\mathsf{e}_i \otimes \mathsf{e}_i}{\mathbf{A}_2\mathsf{e}_i \cdot \mathsf{e}_i}. \tag{1.22}$$

The overall volume fraction of the phase $\mathbf{A}_1$ in the mixture $\mathbf{B}_p$ is

$$\theta = \prod_{i=1}^{p}\theta_i.$$

If this volume fraction θ is fixed, one may wonder what sequential laminates of rank-p can be obtained by varying the proportions θ_i, $i = 1, \ldots, p$. This is given by the following lemma.

Lemma 1.8 Let $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_p$ be unit vectors, $\theta \in [0,1]$ and $m_1, m_2, \ldots, m_p$ nonnegative numbers satisfying

$$\sum_{i=1}^{p} m_i = 1.$$

Then there exists a rank-p sequential laminate $\mathbf{A}$ with matrix $\mathbf{A}_2$ and core $\mathbf{A}_1$, in proportions $(1-\theta)$ and θ, respectively, and lamination directions $\mathsf{e}_1, \mathsf{e}_2 \ldots, \mathsf{e}_p$, such that

$$\theta(\mathbf{A} - \mathbf{A}_2)^{-1} = (\mathbf{A}_1 - \mathbf{A}_2)^{-1} + (1-\theta)\sum_{i=1}^{p} m_i \frac{\mathsf{e}_i \otimes \mathsf{e}_i}{\mathbf{A}_2\mathsf{e}_i \cdot \mathsf{e}_i}. \tag{1.23}$$

The numbers m_i, $i = 1, \ldots, p$ are called the lamination parameters.

Proof. Comparing formulas (1.22) and (1.23), the assertion of lemma is true if there exist $\theta_1, \theta_2, \ldots, \theta_p \in [0,1]$, such that $\theta = \prod_{i=1}^{p}\theta_i$, and

$$(1-\theta)m_i = (1-\theta_i)\prod_{j=1}^{i-1}\theta_j, \quad i = 1, 2, \ldots, p. \tag{1.24}$$

Using formula (1.24) we determine θ_i, $i = 1, \ldots, p$. Since $\sum_{i=1}^{p} m_i = 1$, an easy calculation gives $\prod_{i=1}^{p} \theta_i = \theta$, while from $\theta, m_i \in [0,1]$ it follows $\theta_i \in [0,1]$, $i = 1, \ldots, p$. □

The iterative proces of lamination can be also carried out by using material $\mathbf{A}_1$, instead of $\mathbf{A}_2$ in each lamination. Proceeding analogously, we come to the following result.

Lemma 1.9 Let $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_p$ be unit vectors, $\theta \in [0,1]$ and $m_1, m_2, \ldots, m_p$ nonnegative numbers satisfying

$$\sum_{i=1}^{p} m_i = 1.$$

Then there exists a rank-p sequential laminate $\mathbf{A}$ with matrix $\mathbf{A}_1$ and core $\mathbf{A}_2$, in proportions θ and $(1-\theta)$, respectively, and lamination directions $\mathsf{e}_1, \mathsf{e}_2 \ldots, \mathsf{e}_p$, such that

$$(1-\theta)(\mathbf{A} - \mathbf{A}_1)^{-1} = (\mathbf{A}_2 - \mathbf{A}_1)^{-1} + \theta \sum_{i=1}^{p} m_i \frac{\mathsf{e}_i \otimes \mathsf{e}_i}{\mathbf{A}_1 \mathsf{e}_i \cdot \mathsf{e}_i}. \tag{1.25}$$

□

As will be seen later, the class of sequential laminates appear as optimal composites in optimal design problems. Their explicit representation enables development of various numerical methods for optimal design problems, and thus gives them even more importance. Moreover, sequential laminates saturate optimal bounds on the homogenized matrix $\mathbf{A} \in G_\theta$, expressed through the sum of energies $\sum_{i=1}^{d} \mathbf{A}\boldsymbol{\xi}_i \cdot \boldsymbol{\xi}_i$, for $\boldsymbol{\xi}_i \in \mathbf{R}^d$, $i = 1, \ldots, d$. This sum can be rewritten in the following manner

$$\sum_{i=1}^{d} \mathbf{A}\boldsymbol{\xi}_i \cdot \boldsymbol{\xi}_i = \mathbf{A} : \left(\sum_{i=1}^{d} \boldsymbol{\xi}_i \otimes \boldsymbol{\xi}_i \right) = \mathbf{A} : \boldsymbol{\xi}^\top \boldsymbol{\xi} = (\boldsymbol{\xi}\mathbf{A}) : \boldsymbol{\xi},$$

where $\boldsymbol{\xi} \in \mathrm{M}_d(\mathbf{R})$ is a matrix with rows $\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top, \ldots, \boldsymbol{\xi}_d^\top$, and symbol $:$ stands for the matrix inner product. In order to derive optimal bounds on the G-closure set, we assume that phases $\mathbf{A}_1$ and $\mathbf{A}_2$ are well-ordered, namely,

$$\mathbf{A}_1 \mathsf{e} \cdot \mathsf{e} \leq \mathbf{A}_2 \mathsf{e} \cdot \mathsf{e}, \quad \mathsf{e} \in \mathbf{R}^d.$$

For the non well-ordered case see, e.g. Allaire & Kohn (1993a) and references therein.

Definition 1.3 Let $\theta \in [0,1]$ be the volume fraction of the phase $\mathbf{A}_1$ and $(1-\theta)$ be that of phase $\mathbf{A}_2$, and $\boldsymbol{\xi} = [\boldsymbol{\xi}_1\, \boldsymbol{\xi}_2 \ldots \boldsymbol{\xi}_d]^\top \in \mathrm{M}_d(\mathbf{R})$. A real-valued function $f^+(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$ (respectively, $f^-(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$) is called an upper bound (respectively, a lower bound) if, for any homogenized matrix $\mathbf{A} \in G_\theta$,

$$\mathbf{A} : \boldsymbol{\xi}^\top \boldsymbol{\xi} \leq f^+(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi}) \quad \text{(respectively, } \mathbf{A} : \boldsymbol{\xi}^\top \boldsymbol{\xi} \geq f^-(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})\text{)}.$$

The upper bound $f^{+}(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$ (respectively, lower bound $f^{-}(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$) is said to be optimal if, for any $\boldsymbol{\xi} \in \mathrm{M}_d(\mathbf{R})$, there exists $\mathbf{A}_{\boldsymbol{\xi}} \in G_\theta$ such that

$$\mathbf{A}_\xi : \boldsymbol{\xi}^\top \boldsymbol{\xi} = f^{+}(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi}) \quad (\text{respectively, } \mathbf{A}_\xi : \boldsymbol{\xi}^\top \boldsymbol{\xi} = f^{-}(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})).$$

We already have bounds on the G-closure set from Remark 1.1(v). It is bounded below by the harmonic mean and above by the aritmetic mean,

$$(\theta \mathbf{A}_1^{-1} + (1-\theta)\mathbf{A}_2^{-1})^{-1} : \boldsymbol{\xi}^\top \boldsymbol{\xi} \leq \mathbf{A} : \boldsymbol{\xi}^\top \boldsymbol{\xi} \leq (\theta \mathbf{A}_1 + (1-\theta)\mathbf{A}_2) : \boldsymbol{\xi}^\top \boldsymbol{\xi}, \tag{1.26}$$

almost everywhere on Ω, but these bounds are not optimal (Zhikov et al. (1994)). The optimal bounds, known as the *Hashin-Shtrikman bounds* are given in the following proposition (Hashin & Shtrikman (1963), see also Milton & Kohn (1988), Murat & Tartar (1985), Allaire (2002)).

Proposition 1.10 For any $\boldsymbol{\xi} \in \mathrm{M}_d(\mathbf{R})$, each homogenized matrix $\mathbf{A} \in G_\theta$ satisfies

$$\mathbf{A} : \boldsymbol{\xi}^\top \boldsymbol{\xi} \leq \mathbf{A}_2 : \boldsymbol{\xi}^\top \boldsymbol{\xi} + \theta \min_{\boldsymbol{\eta} \in \mathrm{M}_d(\mathbf{R})} \left[2\boldsymbol{\xi} : \boldsymbol{\eta} + (\mathbf{A}_2 - \mathbf{A}_1)^{-1} : \boldsymbol{\eta}^\top \boldsymbol{\eta} - (1-\theta) h(\boldsymbol{\eta}) \right], \tag{1.27}$$

where $h(\boldsymbol{\eta})$ is a so-called nonlocal term defined by

$$h(\boldsymbol{\eta}) = \min_{\substack{\mathsf{k} \in \mathbf{Z}^d \\ \mathsf{k} \neq 0}} \frac{|\boldsymbol{\eta} \mathsf{k}|^2}{\mathbf{A}_2 \mathsf{k} \cdot \mathsf{k}}$$

and

$$\mathbf{A} : \boldsymbol{\xi}^\top \boldsymbol{\xi} \geq \mathbf{A}_1 : \boldsymbol{\xi}^\top \boldsymbol{\xi} + (1-\theta) \max_{\boldsymbol{\eta} \in \mathrm{M}_d(\mathbf{R})} \left[2\boldsymbol{\xi} : \boldsymbol{\eta} + (\mathbf{A}_2 - \mathbf{A}_1)^{-1} : \boldsymbol{\eta}^\top \boldsymbol{\eta} - \theta g(\boldsymbol{\eta}) \right], \tag{1.28}$$

where $g(\boldsymbol{\eta})$ is a so-called nonlocal term defined by

$$g(\boldsymbol{\eta}) = \max_{\substack{\mathsf{k} \in \mathbf{Z}^d \\ \mathsf{k} \neq 0}} \frac{|\boldsymbol{\eta} \mathsf{k}|^2}{\mathbf{A}_1 \mathsf{k} \cdot \mathsf{k}}.$$

Furthermore, these upper and lower bounds are optimal in the sense of Definition 1.3, and optimality can always be achieved by a rank-d sequential laminate.

□

As stated in the above proposition, Hashin-Shtrikman bounds can be achieved by sequential laminates, but they are not the only optimal microstructures (Tartar (1985), Vigdergauz (1994), Grabovsky (1996)). Furthermore, bounds (1.27) and (1.28) can be equivalently written in the following manner (Allaire 2002, Corollary 2.2.8).

Corollary 1.11 Matrix $\mathbf{A} \in G_\theta$ satisfies the upper Hashin-Shtrikman bound (1.27) if and only if it satisfies

$$\theta(\mathbf{A}_2 - \mathbf{A})^{-1} : \boldsymbol{\eta}^\top \boldsymbol{\eta} \leq (\mathbf{A}_2 - \mathbf{A}_1)^{-1} : \boldsymbol{\eta}^\top \boldsymbol{\eta} - (1-\theta) h(\boldsymbol{\eta}), \quad \boldsymbol{\eta} \in \mathrm{M}_d(\mathbf{R}). \tag{1.29}$$

Moreover, $\mathbf{A} \in G_\theta$ satisfies the lower Hashin-Shtrikman bound (1.28) if and only if it satisfies

$$(1-\theta)(\mathbf{A} - \mathbf{A}_1)^{-1} : \boldsymbol{\eta}^\top \boldsymbol{\eta} \leq (\mathbf{A}_2 - \mathbf{A}_1)^{-1} : \boldsymbol{\eta}^\top \boldsymbol{\eta} - \theta g(\boldsymbol{\eta}), \quad \boldsymbol{\eta} \in \mathrm{M}_d(\mathbf{R}). \tag{1.30}$$

$\square$

Example 1.3 Assume that materials $\mathbf{A}_1$ and $\mathbf{A}_2$ are isotropic, with

$$\mathbf{A}_1 = \alpha \mathbf{I}, \quad \mathbf{A}_2 = \beta \mathbf{I}, \quad 0 < \alpha < \beta.$$

Taking $\boldsymbol{\eta} = \mathbf{I}$, the upper Hashin-Shtrikman bound (1.29) becomes

$$\theta \,\mathrm{tr}\left((\beta \mathbf{I} - \mathbf{A})^{-1}\right) \leq \frac{d}{\beta - \alpha} - \frac{1-\theta}{\beta}, \tag{1.31}$$

while the lower bound (1.30) reads

$$(1-\theta)\,\mathrm{tr}\left((\mathbf{A} - \alpha \mathbf{I})^{-1}\right) \leq \frac{d}{\beta - \alpha} - \frac{\theta}{\alpha}. \tag{1.32}$$

A trace of a matrix is equal to the sum of its eigenvalues, and a simple rearrangement in the two above inequalities leads to

$$\sum_{j=1}^{d} \frac{1}{\lambda_j - \alpha} \leq \frac{1}{\lambda_\theta^- - \alpha} + \frac{d-1}{\lambda_\theta^+ - \alpha}\,, \tag{1.33}$$

$$\sum_{j=1}^{d} \frac{1}{\beta - \lambda_j} \leq \frac{1}{\beta - \lambda_\theta^-} + \frac{d-1}{\beta - \lambda_\theta^+}\,, \tag{1.34}$$

where $\lambda_1, \ldots, \lambda_d$ are eigenvalues of the matrix $\mathbf{A}$, while λ_θ^- and λ_θ^+ are the harmonic and arithmetic means of α and β defined by

$$\lambda_\theta^- = \left(\frac{\theta}{\alpha} + \frac{1-\theta}{\beta}\right)^{-1},$$
$$\lambda_\theta^+ = \theta\alpha + (1-\theta)\beta,$$

as before. Moreover, the bounds (1.26) for isotropic materials read

$$\lambda_\theta^- \leq \lambda_j \leq \lambda_\theta^+\,, \quad j = 1, \ldots, d\,. \tag{1.35}$$

Example 1.3 implies that the set G_θ in the case when mixing two isotropic materials is included in the set of all symmetric matrices whose eigenvalues satisfy (1.33)–(1.35), denoted by $\mathcal{K}(\theta)$. Even more, the G-closure set can be explicitly characterized by this set, which is stated in Theorem 1.13. The set $\Lambda(\alpha, \beta, \theta)$ of all d-tuples $(\lambda_1, \lambda_2, \dots, \lambda_d)$ satisfying (1.33)–(1.35) is given in Figure 1.1 for two and three space dimensions. It can be proven that parts of the boundary of $\Lambda(\alpha, \beta, \theta)$ correspond to sequential laminates in $\mathcal{K}(\theta)$ (see proof of Theorem 1.13). For example, in dimension $d = 2$, points $(\lambda_\theta^-, \lambda_\theta^+)$ and $(\lambda_\theta^+, \lambda_\theta^-)$ in Figure 1.1 correspond to simple laminates, the blue part of the boundary (where equality in (1.34) is achieved) corresponds to rank-2 sequential laminates with core $\mathbf{A}_1$ and matrix $\mathbf{A}_2$, while the red part of the boundary (where equality in (1.33) is achieved) corresponds to rank-2 sequential laminates with core $\mathbf{A}_2$ and matrix $\mathbf{A}_1$. Similarly, when $d = 3$, points denoted in Figure 1.1 correspond to simple laminate, the intersection of the blue surface with one of the planes $\lambda_1 = \lambda_\theta^+$, $\lambda_2 = \lambda_\theta^+$ or $\lambda_3 = \lambda_\theta^+$ corresponds to a rank-2 laminates with core $\mathbf{A}_1$ and matrix $\mathbf{A}_2$, while the rest of the blue surface corresponds to a rank-3 laminates with core $\mathbf{A}_1$ and matrix $\mathbf{A}_2$. Moreover, the intersection of the back surface with one of the planes $\lambda_1 = \lambda_\theta^+$, $\lambda_2 = \lambda_\theta^+$ or $\lambda_3 = \lambda_\theta^+$ corresponds to a rank-2 laminates with core $\mathbf{A}_2$ and matrix $\mathbf{A}_1$, while the rest of the back surface corresponds to a rank-3 laminates with core $\mathbf{A}_2$ and matrix $\mathbf{A}_1$.

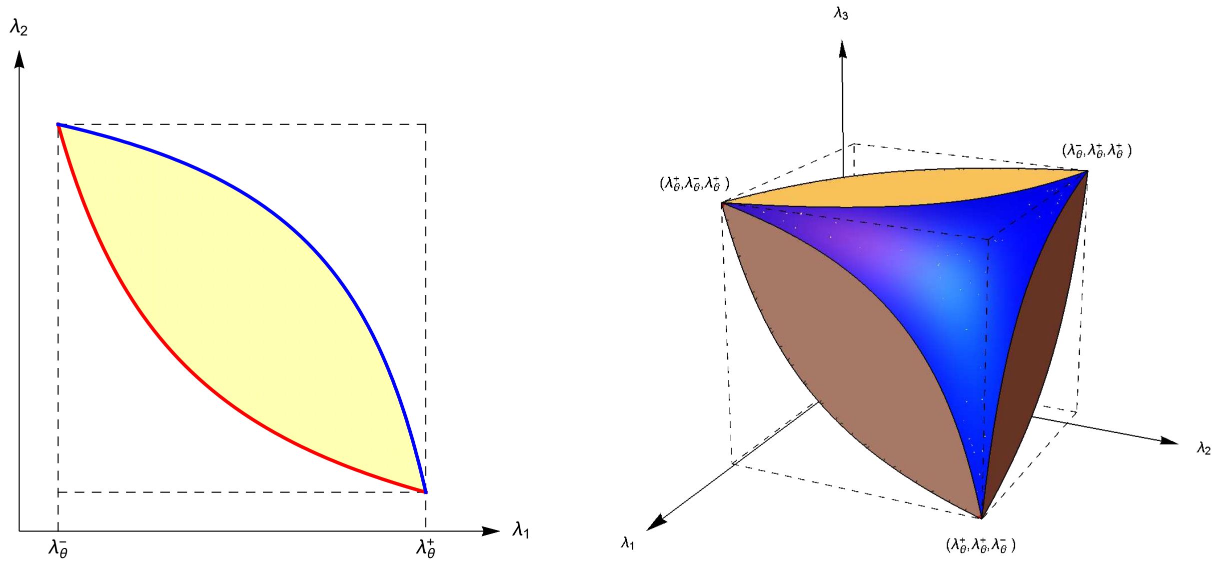


Figure 1.1: Set $\Lambda(\alpha, \beta; \theta)$ in dimensions $d = 2$ and $d = 3$.

Remark 1.3. The bound $\lambda_j \geq \lambda_\theta^-$, in (1.35) can be omitted in the characterization of the set $\mathcal{K}(\theta)$. Indeed, from the inequality $\lambda_j \leq \lambda_\theta^+$, it follows

$$\sum_{j=1}^{d-1} \frac{1}{\lambda_j - \alpha} \geq \sum_{j=1}^{d-1} \frac{1}{\lambda_\theta^+ - \alpha}, \tag{1.36}$$

and together with (1.33) we get $\lambda_j \geq \lambda_\theta^-$. Moreover, if $\mathbf{A}$ has one eigenvalue equal to λ_θ^-, then (1.36) implies equality in (1.33), and all others eigenvalues are equal to λ_θ^+.

An equivalent characterization of $\mathcal{K}(\theta)$ is obtained by considering the eigenvalues of the inverse of tensor $\mathbf{A}$, as stated in lemma below.

Lemma 1.12 The condition $\mathbf{A} \in \mathcal{K}(\theta)$ can be equivalently expressed as $\mathbf{A}^{-1} \in \tilde{\mathcal{K}}(\theta)$, where $\tilde{\mathcal{K}}(\theta)$ is the set of all matrices with eigenvalues ν_j satisfying

$$\nu_\theta^+ \leq \nu_j \leq \nu_\theta^- \,, \quad j = 1, \dots, d \,, \tag{1.37}$$

$$\sum_{j=1}^{d} \frac{1}{\alpha^{-1} - \nu_j} \leq \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{d-1}{\alpha^{-1} - \nu_\theta^+} \,, \tag{1.38}$$

$$\sum_{j=1}^{d} \frac{1}{\nu_j - \beta^{-1}} \leq \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{d-1}{\nu_\theta^+ - \beta^{-1}} \,, \tag{1.39}$$

for $\nu_\theta^+ = \frac{1}{\lambda_\theta^+}$ and $\nu_\theta^- = \frac{1}{\lambda_\theta^-}$. As before, inequalities $\nu_j \leq \nu_\theta^-$ can be omitted.

Proof. Let $\mathbf{A} \in \mathcal{K}(\theta)$, and let us prove that $\nu_j = \frac{1}{\lambda_j}$, satisfy (1.37)–(1.39). From (1.35) it is obvious that (1.37) is valid. Moreover, using (1.33), it follows

$$\begin{aligned}
\sum_{j=1}^{d} \frac{1}{\alpha^{-1} - \nu_j} &= \sum_{j=1}^{d} \frac{\alpha\lambda_j}{\lambda_j - \alpha} = \alpha d + \alpha^2 \sum_{j=1}^{d} \frac{1}{\lambda_j - \alpha} \leq \alpha d + \alpha^2 \left(\frac{1}{\lambda_\theta^- - \alpha} + \frac{d-1}{\lambda_\theta^+ - \alpha} \right) = \\
&= \alpha \left(\frac{\lambda_\theta^-}{\lambda_\theta^- - \alpha} + \frac{(d-1)\lambda_\theta^+}{\lambda_\theta^+ - \alpha} \right) = \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{d-1}{\alpha^{-1} - \nu_\theta^+},
\end{aligned}$$

while from (1.34) it follows

$$\begin{aligned}
\sum_{j=1}^{d} \frac{1}{\nu_j - \beta^{-1}} &= \sum_{j=1}^{d} \frac{\beta\lambda_j}{\beta - \lambda_j} = \beta^2 \sum_{j=1}^{d} \frac{1}{\beta - \lambda_j} - \beta d \leq \beta^2 \left(\frac{1}{\beta - \lambda_\theta^-} + \frac{d-1}{\beta - \lambda_\theta^+} \right) - \beta d = \\
&= \beta \left(\frac{\lambda_\theta^-}{\beta - \lambda_\theta^-} + \frac{(d-1)\lambda_\theta^+}{\beta - \lambda_\theta^+} \right) = \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{d-1}{\nu_\theta^+ - \beta^{-1}}.
\end{aligned}$$

Therefore, $\mathbf{A}^{-1} \in \tilde{\mathcal{K}}(\theta)$. Reverse inclusion can be proven analogously. □

Remark 1.4. Let us denote the set of all $(\nu_1, \dots, \nu_d)$ satisfying (1.37)–(1.39) by $\mathcal{V}(\alpha, \beta; \theta)$, and $\Lambda(\alpha, \beta; \theta)$ as before. Note that in the case when $d = 2$, the set $\mathcal{V}(\alpha, \beta; \theta)$ is equal to the set $\Lambda\left(\frac{1}{\beta}, \frac{1}{\alpha}; 1 - \theta\right)$. This is not true in general. For example, take a simple laminate with $\nu = (\nu_\theta^-, \nu_\theta^+, \dots, \nu_\theta^+) \in \mathcal{V}(\alpha, \beta; \theta)$ (in (1.38) and (1.39) equalities are achieved). For ν_j, $j = 1, \dots, d$ to be in $\Lambda\left(\frac{1}{\beta}, \frac{1}{\alpha}; 1 - \theta\right)$, it must satisfy (1.33)–(1.35), i.e. the following inequalities must be satisfied:

$$\nu_\theta^+ \leq \quad \nu_j \quad \leq \nu_\theta^- \,, \quad j = 1, \ldots, d \,, \tag{1.40}$$

$$\sum_{j=1}^{d} \frac{1}{\nu_j - \beta^{-1}} \quad \leq \quad \frac{1}{\nu_\theta^+ - \beta^{-1}} + \frac{d-1}{\nu_\theta^- - \beta^{-1}} \,, \tag{1.41}$$

$$\sum_{j=1}^{d} \frac{1}{\alpha^{-1} - \nu_j} \quad \leq \quad \frac{1}{\alpha^{-1} - \nu_\theta^+} + \frac{d-1}{\alpha^{-1} - \nu_\theta^-} \,. \tag{1.42}$$

This is not possible, except for $d = 2$, since for $d > 2$, from (1.41) and equality in (1.38), it follows $\nu_\theta^+ \geq \nu_\theta^-$.

Remark 1.5. Sets $\tilde{\mathcal{K}}(\theta)$ and $\mathcal{K}(\theta)$, are convex subsets of the set of symmetric positive definite matrices. Indeed, functions $f_1(\nu_1, \ldots, \nu_d) = \sum_{i=1}^{d} \frac{1}{\nu_i - \beta^{-1}}$ and $f_2(\nu_1, \ldots, \nu_d) = \sum_{i=1}^{d} \frac{1}{\alpha^{-1} - \nu_i}$ are convex functions on $[\nu_\theta^+, \nu_\theta^-]^d$. Since a convex function of the eigenvalues is also convex function with respect to the associated symmetric positive definite matrices (Ball (1977)), functions $F_1(\mathbf{A}) := f_1(\nu_1, \ldots, \nu_d)$ and $F_2(\mathbf{A}) := f_2(\nu_1, \ldots, \nu_d)$ are convex functions on the set of all symmetric positive definite matrices, denoted by Sym_d^+. Here, $\nu_1, \ldots, \nu_d$ are eigenvalues of the matrix $\mathbf{A} \in \mathrm{Sym}_d^+$ and they belong to $[\nu_\theta^+, \nu_\theta^-]^d$. Moreover, the sublevel set of a convex function is a convex set, which implies that sets

$$\left\{ \mathbf{A} \in \mathrm{Sym}_d^+ : F_1(\mathbf{A}) \leq \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{d-1}{\alpha^{-1} - \nu_\theta^+} \right\}$$

and

$$\left\{ \mathbf{A} \in \mathrm{Sym}_d^+ : F_2(\mathbf{A}) \leq \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{d-1}{\nu_\theta^+ - \beta^{-1}} \right\}$$

are convex sets. Therefore, the set $\tilde{\mathcal{K}}(\theta)$ is convex, as the intersection of convex sets. Similarly, convexity of the set $\mathcal{K}(\theta)$ follows from the convexity of functions $\sum_{i=1}^{d} \frac{1}{\lambda_i - \alpha}$ and $\sum_{i=1}^{d} \frac{1}{\beta - \lambda_i}$ on $[\lambda_\theta^-, \lambda_\theta^+]^d$.

Theorem 1.13 For $\theta \in [0, 1]$, $\mathbf{A}_1 = \alpha \mathbf{I}$, $\mathbf{A}_2 = \beta \mathbf{I}$, $0 < \alpha < \beta$, it follows $G_\theta = \mathcal{K}(\theta)$.

For the proof of the above theorem, we shall need the following lemma (Allaire 2002, Lemma 2.2.9).

Lemma 1.14 Denote by G_θ^d the G-closure set in d space dimensions. If $\tilde{\mathbf{A}} \in G_\theta^{d-1}$, then

$$\begin{bmatrix} \tilde{\mathbf{A}} & 0 \\ 0 & \lambda_\theta^+ \end{bmatrix} \in G_\theta^d.$$

□

Proof of Theorem 1.13. We already proved the inclusion $G_\theta \subseteq \mathcal{K}(\theta)$, and the reverse inclusion we treat by induction on the space dimension d. Let us denote by G_θ^d and $\mathcal{K}^d(\theta)$

spaces G_θ and $\mathcal{K}(\theta)$ in the case of dimension d. For $d = 1$, inequalities (1.33) and (1.34) give $\mathcal{K}^1(\theta) = \{\lambda_\theta^-\}$. Moreover, in this case, from Example 1.1, we know that G_θ^1 reduces to $\{\lambda_\theta^-\}$. Assume now that $\mathcal{K}^{d-1}(\theta) \subseteq G_\theta^{d-1}$. To prove $\mathcal{K}^d(\theta) \subseteq G_\theta^d$, we shall first prove that this inclusion is valid for boundary points of $\mathcal{K}^d(\theta)$. The boundary of $\mathcal{K}(\theta)$ is defined by at least one of inequalities (1.33)–(1.35) being saturated. By Remark 1.3, inequalities $\lambda_j \geq \lambda_\theta^-$ do not need to be considered. First, let one eigenvalue be equal to the arithmetic mean. With no loss of generality, let $\lambda_d = \lambda_\theta^+$. Obviously, $(\lambda_1, \lambda_2, \ldots, \lambda_{d-1})$ satisfy $d-1$ analogous inequalities (1.33)–(1.35), and therefore, a symmetric $(d-1) \times (d-1)$ matrix $\tilde{\mathbf{A}}$ having these eigenvalues belongs to set $\in \mathcal{K}^{d-1}(\theta) \subseteq G_\theta^{d-1}$. Then, Lemma 1.14 implies

$$\mathbf{A} = \begin{bmatrix} \tilde{\mathbf{A}} & 0 \\ 0 & \lambda_\theta^+ \end{bmatrix} \in G_\theta^d.$$

Consider now a symmetric matrix $\mathbf{A} \in \mathcal{K}^d(\theta)$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ satisfying strict inequalities in (1.35) and equality in (1.34), i.e.

$$\sum_{j=1}^{d} \frac{1}{\beta - \lambda_j} = \frac{1}{\beta - \lambda_\theta^-} + \frac{d-1}{\beta - \lambda_\theta^+}. \tag{1.43}$$

Denote by $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_d$ eigenvectors of $\mathbf{A}$ (which form the orthonormal basis), associated to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$. We shall prove that eigenvalues of the matrix $\mathbf{A}$ are also eigenvalues of a rank-d sequential laminate, which belongs to G_θ^d. Consider a rank-d sequential laminate with matrix $\beta\mathbf{I}$ and core $\alpha\mathbf{I}$, lamination directions $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_d$ and parameters $m_i \geq 0$, $i = 1, \ldots, d$, $\sum_{i=1}^d m_i = 1$. From Lemma 1.8, conductivity $\tilde{\mathbf{A}}$ of such laminate is given by

$$\theta(\tilde{\mathbf{A}} - \beta\mathbf{I})^{-1} = \frac{1}{\alpha - \beta}\mathbf{I} + \frac{1-\theta}{\beta} \sum_{i=1}^{d} m_i \mathsf{e}_i \otimes \mathsf{e}_i.$$

Notice that $\tilde{\mathbf{A}}$ has eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ if and only if parameters m_i can be chosen such that

$$\frac{\theta}{\lambda_i - \beta} = \frac{1}{\alpha - \beta} + \frac{1-\theta}{\beta} m_i \quad i = 1, \ldots, d. \tag{1.44}$$

It is easy to check that m_i defined by (1.44) satisfies $0 < m_i < 1$ if and only if $\lambda_\theta^- < \lambda_i < \lambda_\theta^+$. Furthermore, $\sum_{i=1}^d m_i = 1$ if and only if equality (1.43) holds, and therefore, the eigenvalues of the matrix $\mathbf{A}$ are those of a rank-d sequential laminate $\tilde{\mathbf{A}} \in G_\theta^d$. Since $\mathbf{A}$ and $\tilde{\mathbf{A}}$ are symmetric matrices with the same eigenvalues and eigenvectors, it follows $\mathbf{A} = \tilde{\mathbf{A}}$ and $\mathbf{A} \in G_\theta^d$. We can apply an analogous procedure if the eigenvalues of $\mathbf{A}$ satisfy strict inequalities in (1.35), and equality in (1.33), and get $\mathbf{A} \in G_\theta^d$.

It remains to prove that the same is true for the interior points of $\mathcal{K}^d(\theta)$, where none of the equalities in (1.33)–(1.35) is achieved. To do so, we shall prove that any tensor in the interior can be obtained as a simple laminate of two boundary tensors, and, since G_θ^d is stable under lamination, meaning that a laminate obtained from two materials from G_θ^d belongs to G_θ^d, this will conclude the proof. Therefore, assume that strict inequalities in (1.33) and (1.34) are valid for eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_d$ of the matrix $\mathbf{A}$. We keep $\lambda_1, \lambda_2, \ldots, \lambda_{d-1}$ fixed and increase or decrease eigenvalue λ_d until boundary of $\mathcal{K}^d(\theta)$ is met. In this way we obtain two points on the boundary, $(\lambda_1, \ldots, \lambda_{d-1}, \lambda_d^-)$ and $(\lambda_1, \ldots, \lambda_{d-1}, \lambda_d^+)$, such that λ_d is a convex combination of λ_d^- and λ_d^+. Let $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_d$ be an othonormal basis of eigenvectors of $\mathbf{A}$, and define $\mathbf{A}_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_{d-1}, \lambda_d^-)$, $\mathbf{A}_2 = \operatorname{diag}(\lambda_1, \ldots, \lambda_{d-1}, \lambda_d^+)$ in this basis. Applying formula (1.18) for simple laminates to materials $\mathbf{A}_1$ and $\mathbf{A}_2$ with proportions t, $(1-t)$ respectively, and lamination direction e_d yields a homogenized tensor

$$\tilde{\mathbf{A}} = t\mathbf{A}_1 + (1-t)\mathbf{A}_2 - \frac{t(1-t)(\lambda_d^- - \lambda_d^+)^2 \mathsf{e}_d \otimes \mathsf{e}_d}{(1-t)\lambda_d^- + t\lambda_d^+}.$$

Tensor $\tilde{\mathbf{A}}$ has eigenvectors $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_d$ and eigenvalues

$$\begin{aligned} \lambda_i(t) &= \lambda_i, \quad i = 1, \ldots, d-1 \\ \lambda_d(t) &= t\lambda_d^- + (1-t)\lambda_d^+ - \frac{t(1-t)(\lambda_d^- - \lambda_d^+)^2}{(1-t)\lambda_d^- + t\lambda_d^+}. \end{aligned}$$

By choosing an adequate $t \in (0,1)$ we can obtain $\lambda_d(t) = \lambda_d$ and conclude $\mathbf{A} = \tilde{\mathbf{A}} \in G_\theta^d$. □

From Theorem 1.13 and Lemma 1.12 it follows that the set G_θ can be also characterized by the set $\tilde{\mathcal{K}}(\theta)$.

1.3 Homogenization in linearized elasticity

The definition and results on H-convergence from the Section 1.1 can be adapted to the elasticity setting (Francfort & Murat (1986), Tartar (1986), Zhikov et al. (1994), see also Zhikov et al. (1979)). In this case, we consider the elastic domain $\Omega \subseteq \mathbf{R}^d$ whose mechanical behaviour under some force load $\mathsf{f} \in \mathrm{H}^{-1}(\Omega; \mathbf{R}^d)$ are described with the linearized elasticity system

$$\begin{cases} -\operatorname{div}(\mathbf{A}e(\mathsf{u})) = \mathsf{f} \quad \text{in } \Omega \\ \mathsf{u} \in \mathrm{H}_0^1(\Omega; \mathbf{R}^d), \end{cases} \tag{1.45}$$

where $e(\mathsf{u}) = \frac{1}{2}(\nabla \mathsf{u} + \nabla \mathsf{u}^\top)$. The function u represents the displacement uniquely determined by the force density f, while the fourth order tensor $\mathbf{A}$, called the stiffness tensor, contains information about elastic properties of a material. Matrix $e(\mathsf{u})$ is known as the strain tensor, and it describes the deformation of the material, while $\boldsymbol{\sigma} := \mathbf{A}e(\mathsf{u})$ is the

stress tensor which expresses the internal forces that caused the deformation.
Denote by

$$\mathrm{Sym}_d^4 := \{\mathbf{A} = [a_{ijkl}]_{1\le i,j,k,l\le d} \in \mathbf{R}^{d\times d\times d\times d} : a_{ijkl} = a_{jikl} = a_{ijlk} = a_{klij}\}$$

the space of all symmetric fourth order tensors, and for $0 < \alpha < \beta$, define a set

$$\mathcal{M}^4_{\alpha,\beta} := \{\mathbf{A} \in \mathrm{Sym}_d^4 : \mathbf{A}\boldsymbol{\xi} : \boldsymbol{\xi} \ge \alpha|\boldsymbol{\xi}|^2,\ \mathbf{A}^{-1}\boldsymbol{\xi} : \boldsymbol{\xi} \ge \frac{1}{\beta}|\boldsymbol{\xi}|^2,\ \boldsymbol{\xi} \in \mathrm{Sym_d}\}.$$

Tensor $\mathbf{A} \in \mathrm{Sym}_d^4$ can be considered as a linear operator $\mathbf{A} \in \mathcal{L}(\mathrm{Sym}_d, \mathrm{Sym}_d)$, defined with $[\mathbf{A}\boldsymbol{\xi}]_{ij} := \left[\sum_{k,l=1}^d a_{ijkl}\xi_{kl}\right]_{ij}$, for $\boldsymbol{\xi} \in \mathrm{Sym}_d$. The weak formulation of equation (1.45) for $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathcal{M}^4_{\alpha,\beta})$ reads

$$\int_\Omega \mathbf{A}e(\mathsf{u}) : e(\boldsymbol{\varphi})\,d\mathbf{x} = \int_\Omega \mathsf{f}\cdot\boldsymbol{\varphi}\,d\mathbf{x}, \quad \boldsymbol{\varphi} \in \mathrm{H}^1_0(\Omega; \mathbf{R}^d), \tag{1.46}$$

and, since $\mathbf{A}$ is bounded and coercive, the Lax-Milgram lemma implies that the equation admits a unique weak solution.

Definition 1.4 A sequence of tensors $\mathbf{A}^n \in \mathrm{L}^\infty(\Omega; \mathcal{M}^4_{\alpha,\beta})$ is said to H-converge to a homogenized limit, or H-limit, $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathcal{M}^4_{\alpha',\beta'})$ if, for any right hand side $\mathsf{f} \in \mathrm{H}^{-1}(\Omega; \mathbf{R}^d)$, the sequence u^n of solutions of

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^n e(\mathsf{u}^n)) = \mathsf{f} & \text{in } \Omega \\ \mathsf{u}^n \in \mathrm{H}^1_0(\Omega; \mathbf{R}^d) \end{cases}$$

satisfies

$$\begin{cases} \mathsf{u}^n \rightharpoonup \mathsf{u} \text{ in } \mathrm{H}^1_0(\Omega; \mathbf{R}^d) \\ \mathbf{A}^n e(\mathsf{u}^n) \rightharpoonup \mathbf{A}e(\mathsf{u}) \text{ in } \mathrm{L}^2(\Omega; \mathrm{Sym}_d)\,, \end{cases}$$

and consequently, u is the solution of the homogenized equation

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}e(\mathsf{u})) = \mathsf{f} & \text{in } \Omega \\ \mathsf{u} \in \mathrm{H}^1_0(\Omega; \mathbf{R}^d). \end{cases}$$

As in the conductivity case, the above definition makes sense because of the following compactness theorem (Francfort & Murat (1986)).

Theorem 1.15 For any sequence $\mathbf{A}^n \in \mathrm{L}^\infty(\Omega; \mathcal{M}^4_{\alpha,\beta})$ there exists a subsequence which H-converges to $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathcal{M}^4_{\alpha,\beta})$.

□

Likewise, all statements from Remark 1.1 are valid in the elasticity setting as well. We already emphasized a value of laminated composites in the previous section, so let us describe this subset of composites in the elasticity setting. Consider two isotropic elastic phases with stiffness tensors

$$\begin{aligned} \mathbf{A}_1 &= 2\mu_1\mathbf{I}_4 + \left(\kappa_1 - \frac{2\mu_1}{d}\right)\mathbf{I}_2 \otimes \mathbf{I}_2 \\ \mathbf{A}_2 &= 2\mu_2\mathbf{I}_4 + \left(\kappa_2 - \frac{2\mu_2}{d}\right)\mathbf{I}_2 \otimes \mathbf{I}_2, \end{aligned}$$

for μ_1, μ_2 being the shear moduli, and κ_1, κ_2 the bulk moduli, $\mathbf{I}_4$ an identity tensor in Sym_d^4 and $\mathbf{I}_2$ the identity matrix. The terms $\lambda_1 = \kappa_1 - \frac{2\mu_1}{d}$ and $\lambda_2 = \kappa_2 - \frac{2\mu_2}{d}$ are known as Lamé coefficients. We shall assume that the phases $\mathbf{A}_1$ and $\mathbf{A}_2$ are well ordered, i.e.

$$0 < \mu_1 \leq \mu_2, \quad 0 < \kappa_1 \leq \kappa_2.$$

A formula for simple laminate, made of materials $\mathbf{A}_1$ and $\mathbf{A}_2$ with proportion θ of the first material and layers perpendicular to the lamination direction e is given in the following lemma (Francfort & Murat (1986), see also Allaire (2002)).

Lemma 1.16 Let $\mathsf{e} \in \mathbf{R}^d$ be a unit vector and $\chi^n(\mathbf{x} \cdot \mathsf{e})$ a sequence of characteristic functions that weak-$*$ converges to a limit $\theta(\mathbf{x}\cdot\mathsf{e})$ in $\mathrm{L}^\infty(\Omega; [0,1])$. For two isotropic tensors $\mathbf{A}_1$ and $\mathbf{A}_2$ in $\mathcal{M}^4_{\alpha,\beta}$, a sequence of tensors $\mathbf{A}^n(\mathbf{x} \cdot \mathsf{e}) = \chi^n(\mathbf{x} \cdot \mathsf{e})\mathbf{A}_1 + (1 - \chi^n(\mathbf{x} \cdot \mathsf{e}))\mathbf{A}_2$ H-converges to $\mathbf{A}(\mathbf{x} \cdot \mathsf{e})$, which is given by the formula

$$\theta(\mathbf{A} - \mathbf{A}_2)^{-1} = (\mathbf{A}_1 - \mathbf{A}_2)^{-1} + (1-\theta) f_2(\mathsf{e}), \tag{1.47}$$

where $f_2(\mathsf{e})$ is symmetric positive semidefinite fourth order tensor defined by the quadratic form

$$f_2(\mathsf{e})\boldsymbol{\xi} : \boldsymbol{\xi} = \frac{1}{\mu_2}\left(|\boldsymbol{\xi}\mathsf{e}|^2 - (\boldsymbol{\xi}\mathsf{e}\cdot\mathsf{e})^2\right) + \frac{1}{2\mu_2 + \lambda_2}(\boldsymbol{\xi}\mathsf{e}\cdot\mathsf{e})^2, \quad \boldsymbol{\xi} \in \mathrm{Sym_d}. \tag{1.48}$$

Additionally, the tensor $\mathbf{A}$ can equivalently be given by the formula

$$\theta(\mathbf{A}^{-1} - \mathbf{A}_2^{-1})^{-1} = (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1} + (1-\theta) f_2^c(\mathsf{e}), \tag{1.49}$$

where $f_2^c(\mathsf{e})$ is symmetric positive semidefinite fourth order tensor defined by the quadratic form

$$f_2^c(\mathsf{e})\boldsymbol{\xi} : \boldsymbol{\xi} = \mathbf{A}_2\boldsymbol{\xi} : \boldsymbol{\xi} - \frac{1}{\mu_2}|\mathbf{A}_2\boldsymbol{\xi}\mathsf{e}|^2 + \frac{\mu_2 + \lambda_2}{\mu_2(2\mu_2+\lambda_2)}\left((\mathbf{A}_2\boldsymbol{\xi})\mathsf{e}\cdot\mathsf{e}\right)^2, \quad \boldsymbol{\xi} \in \mathrm{Sym_d}. \tag{1.50}$$

□

For different choices for θ and e, formula (1.47) (i.e. formula (1.49)) yields a whole family of single laminates made from phases $\mathbf{A}_1$ and $\mathbf{A}_2$. We can also use formula (1.47) in order to get laminates of a higher rank. For example, if we iterate formula (1.47) p times, and in each iteration use material obtained in the previous one and $\mathbf{A}_2$, then the obtained composite is determined by

$$\left(\prod_{i=1}^{p} \theta_i\right) (\mathbf{A}^p - \mathbf{A}_2)^{-1} = (\mathbf{A}_1 - \mathbf{A}_2)^{-1} + \sum_{i=1}^{p} \left((1-\theta_i) \prod_{j=1}^{i-1} \theta_j\right) f_2(\mathsf{e}_i), \tag{1.51}$$

where e_i, $i = 1, \ldots, p$ are the lamination directions in each iteration, and $f_2(\mathsf{e}_i)$ is given by (1.48). We call it a rank-p sequential laminate with the matrix phase $\mathbf{A}_2$ and core phase $\mathbf{A}_1$. It is characterized by the lamination directions $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_p$ and the proportions $\theta_1, \theta_2, \ldots, \theta_p$ in each stage of the process. The overall volume fraction of phase $\mathbf{A}_1$ in the above formula is $\theta = \prod_{i=1}^{p} \theta_i$. A similar formula can be obtained by using phase $\mathbf{A}_1$ instead of $\mathbf{A}_2$ in each lamination. If a volume fraction θ is fixed, we can get a parametrized family of rank-p sequential laminates (Allaire 2002, Lemma 2.3.3).

Lemma 1.17 Let $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_p$ be a collection of unit vectors and $\theta \in [0, 1]$ be a volume fraction. For any collection of numbers $m_i \geq 0$, $i = 1, \ldots, p$ satisfying $\sum_{i=1}^{p} m_i = 1$ there exists a rank-p sequential laminate $\mathbf{A}$, with core phase $\mathbf{A}_1$ and matrix phase $\mathbf{A}_2$, in proportions θ and $(1-\theta)$, respectively, defined by

$$\theta(\mathbf{A} - \mathbf{A}_2)^{-1} = (\mathbf{A}_1 - \mathbf{A}_2)^{-1} + (1-\theta) \sum_{i=1}^{p} m_i f_2(\mathsf{e}_i), \tag{1.52}$$

where f_2 is given by (1.48). The numbers m_i, $i = 1, \ldots, p$ are called the lamination paramaters.

By interchanging the roles of $\mathbf{A}_1$ and $\mathbf{A}_2$, there exists a rank-p sequential laminate $\mathbf{B}$, with matrix phase $\mathbf{A}_1$ and core phase $\mathbf{A}_2$, in proportions θ and $(1-\theta)$, respectively, with lamination directions $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_p$ such that

$$(1-\theta)(\mathbf{B} - \mathbf{A}_1)^{-1} = (\mathbf{A}_2 - \mathbf{A}_1)^{-1} + \theta \sum_{i=1}^{p} m_i f_1(\mathsf{e}_i), \tag{1.53}$$

where $f_1(\mathsf{e}_i)$ is given by

$$f_1(\mathsf{e})\boldsymbol{\xi} : \boldsymbol{\xi} = \frac{1}{\mu_1}\left(|\boldsymbol{\xi}\mathsf{e}|^2 - (\boldsymbol{\xi}\mathsf{e}\cdot\mathsf{e})^2\right) + \frac{1}{2\mu_1 + \lambda_1}(\boldsymbol{\xi}\mathsf{e}\cdot\mathsf{e})^2, \quad \boldsymbol{\xi} \in \mathrm{Sym}_{\mathrm{d}}. \tag{1.54}$$

□

From the formula (1.52), if the lamination parameters $m_1, \ldots, m_p$ and density θ are known, one can get proportions $\theta_1, \ldots, \theta_p$ of the first phase in each lamination from the formula

$$(1-\theta)m_i = (1-\theta_i) \prod_{j=1}^{i-1} \theta_j, \quad i = 1, \ldots, p. \tag{1.55}$$

As before, formulas for sequential laminates can be equivalently expressed in terms of inverse tensors $\mathbf{A}_1^{-1}$ and $\mathbf{A}_2^{-1}$, also called *compliance tensors*. Thus the formula (1.52) is equivalent to

$$\theta(\mathbf{A}^{-1} - \mathbf{A}_2^{-1})^{-1} = (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1} + (1-\theta)\sum_{i=1}^{p} m_i f_2^c(\mathsf{e}_i), \tag{1.56}$$

where $f_2^c(\mathsf{e}_i)$ is given by (1.50), $i = 1, \ldots, p$ and formula (1.53) is equivalent to

$$(1-\theta)(\mathbf{B}^{-1} - \mathbf{A}_1^{-1})^{-1} = (\mathbf{A}_2^{-1} - \mathbf{A}_1^{-1})^{-1} + \theta\sum_{i=1}^{p} m_i f_1^c(\mathsf{e}_i), \tag{1.57}$$

where $f_1^c(\mathsf{e}_i)$, $i = 1, \ldots, p$ is given by

$$f_1^c(\mathsf{e}_i)\boldsymbol{\xi} : \boldsymbol{\xi} = \mathbf{A}_1\boldsymbol{\xi} : \boldsymbol{\xi} - \frac{1}{\mu_1}|\mathbf{A}_1\boldsymbol{\xi}\mathsf{e}_i|^2 + \frac{\mu_1 + \lambda_1}{\mu_1(2\mu_1 + \lambda_1)}\left((\mathbf{A}_1\boldsymbol{\xi})\mathsf{e}_i \cdot \mathsf{e}_i\right)^2, \quad \boldsymbol{\xi} \in \mathrm{Sym_d}. \tag{1.58}$$

This representation of a sequential laminate in terms of compliance tensors is useful when working with stresses rather than strains, as will be seen later.

Definition 1.5 For $\theta \in [0,1]$, we denote by L_θ^+ the set of all sequential laminates $\mathbf{A}$, with core phase $\mathbf{A}_1$ and matrix phase $\mathbf{A}_2$, in proportions θ and $(1-\theta)$, respectively, defined by formula (1.52) (or, equivalently, (1.56)), obtained by varying the number p of laminations and the lamination directions $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_p$, as well as the lamination paremeters $m_1, \ldots, m_p$.

Moreover, we denote by L_θ^- the set of all sequential laminates $\mathbf{B}$, with matrix phase $\mathbf{A}_1$ and core phase $\mathbf{A}_2$, in proportions θ and $(1-\theta)$, respectively, defined by formula (1.53) (or, equivalently, (1.57)), obtained by varying the number p of laminations and the lamination directions $\mathsf{e}_1, \mathsf{e}_2, \ldots, \mathsf{e}_p$, as well as the lamination paremeters $m_1, \ldots, m_p$.

There exists a parametrization of the sets L_θ^+ and L_θ^- that we shall use later, hence we state it in the following lemma (Avellaneda (1987), see also Allaire (2002)),

Lemma 1.18 The set L_θ^+ is the bounded closed subset of all symmetric fourth order tensors $\mathbf{A} \in \mathrm{Sym}_d^4$ such that there exists a probability measure ν on the unit sphere S^{d-1} satisfying

$$\theta(\mathbf{A}^{-1} - \mathbf{A}_2^{-1})^{-1} = (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1} + (1-\theta)\int_{S^{d-1}} f_2^c(\mathsf{e})\, d\nu(\mathsf{e}), \tag{1.59}$$

where $f_2^c(\mathsf{e})$ is given by (1.50). Furthermore, any tensor $\mathbf{A} \in L_\theta^+$ is a tensor of a finite-rank sequential laminate defined by (1.56) with rank $p \leq \frac{1}{24}d(d+1)(d+2)(d+3)$.

Similarly, L_θ^- is the bounded closed subset of all symmetric fourth order tensors $\mathbf{B} \in \mathrm{Sym}_d^4$ such that there exists a probability measure ν on the unit sphere S^{d-1} satisfying

$$(1-\theta)(\mathbf{B}^{-1} - \mathbf{A}_1^{-1})^{-1} = (\mathbf{A}_2^{-1} - \mathbf{A}_1^{-1})^{-1} + \theta\int_{S^{d-1}} f_1^c(\mathsf{e})\, d\nu(\mathsf{e}), \tag{1.60}$$

where $f_1^c(\mathsf{e})$ is given by (1.58). Furthermore, any tensor $\mathbf{A} \in L_\theta^-$ is a tensor of a finite-rank sequential laminate defined by (1.57) with rank $p \leq \frac{1}{24}d(d+1)(d+2)(d+3)$.

$\square$

Let us now try to describe the G-closure set in the elasticity setting, i.e. the set of all composite materials obtained by mixing two isotropic elastic phases $\mathbf{A}_1$ and $\mathbf{A}_2$ in proportions θ and $(1-\theta)$, respectively. Theorem 1.5 is valid in elasticity setting as well, meaning that we have the locality property of the G-closure, and we shall describe the set G_θ in the sequel. Unlike the conductivity case, where we have an explicit characterization of the G-closure set for mixing two isotropic materials, in the linearized elasticity the set is still unknown. All we can do is to derive optimal bounds on homogenized elasticity tensor, i.e. Hashin-Shtrikman bounds. Before providing the bounds, let us recall the definition of the optimal bounds on elasticity tensors $\mathbf{A} \in G_\theta$.

Definition 1.6 Let $\theta \in [0,1]$ be the volume fraction of the phase $\mathbf{A}_1$ and $(1-\theta)$ be that of phase $\mathbf{A}_2$, and $\boldsymbol{\xi} \in \mathrm{Sym}_d$. A real-valued function $f^+(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$ (respectively, $f^-(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$) is called an upper bound (respectively, a lower bound) if, for any homogenized matrix $\mathbf{A} \in G_\theta$,

$$\mathbf{A}\boldsymbol{\xi} : \boldsymbol{\xi} \leq f^+(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi}) \quad \text{(respectively, } \mathbf{A}\boldsymbol{\xi} : \boldsymbol{\xi} \geq f^-(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})\text{)}.$$

The upper bound $f^+(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$ (respectively, lower bound $f^-(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})$) is said to be optimal if, for any $\boldsymbol{\xi} \in \mathrm{M}_d(\mathbf{R})$, there exists $\mathbf{A}_{\boldsymbol{\xi}} \in G_\theta$ such that

$$\mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\xi} : \boldsymbol{\xi} = f^+(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi}) \quad \text{(respectively, } \mathbf{A}_{\boldsymbol{\xi}}\boldsymbol{\xi} : \boldsymbol{\xi} = f^-(\theta, \mathbf{A}_1, \mathbf{A}_2; \boldsymbol{\xi})\text{)}.$$

The above definition can easily be extended to the sum of energies, $\sum_{i=1}^p \mathbf{A}\boldsymbol{\xi}_i : \boldsymbol{\xi}_i$, for symmetric matrices $\boldsymbol{\xi}_i$, $i = 1, \dots p$, and the optimal bounds (Hashin-Shtrikman bounds) are known in this case as well. We shall focus on the single energy bounds, and for the multiple case we refer to Allaire (2002). The bounds on the elastic energy can be written in terms of strain (primal energy), as bounds on $\mathbf{A}\boldsymbol{\xi} : \boldsymbol{\xi}$, or in terms of stress (complementary energy), as bounds on $\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma}$. We give the lower and the upper bound on the complementary energy in Proposition 1.21. In order to prove the proposition, we shall need some additional results, which we give in the sequel.

Definition 1.7 Let E be a normed vector space and $\varphi : E \to \langle -\infty, \infty]$ a function such that $\varphi \not\equiv +\infty$. We define the conjugate function $\varphi^* : E' \to \langle -\infty, \infty]$ to be

$$\varphi^*(f) = \sup_{x \in E} \left[{}_{E'}\langle f, x \rangle_E - \varphi(x) \right], \quad f \in E'.$$

Function φ^* is called the Legendre transform of φ.

Note that φ^* is convex and lower semicontinuous on E'. We also define

$$\varphi^{**}(x) = \sup_{f \in E'} \left[{}_{E'}\langle f, x \rangle_E - \varphi^*(f) \right], \quad x \in E.$$

Obviously, $\varphi^{**} \leq \varphi$, while the equality is valid in some special case, as stated in theorem below (Brezis 2010, Theorem 1.11).

Theorem 1.19 (Fenchel-Moreau) Assume that $\varphi : E \to \langle -\infty, \infty]$ is convex, lower semicontinuous and $\varphi \not\equiv \infty$. Then $\varphi^{**} = \varphi$.

□

We shall also need the following minimax theorem (Aubin & Ekeland 1984, Chapter 6).

Proposition 1.20 Let K be a compact convex subset of a topological vector space, let U be a convex subset of a vector space, and let $h : K \times U \to \mathbf{R}$ be a function such that:

(i) For each $u \in U$, the function $x \mapsto h(x, u)$ is convex and lower semicontinuous,

(ii) For each $x \in K$, the function $u \mapsto -h(x, u)$ is convex.

Then

$$\inf_{x \in K} \sup_{u \in U} h(x, u) = \sup_{u \in U} \inf_{x \in K} h(x, y).$$

□

Proposition 1.21 Let $\boldsymbol{\sigma} \in \mathrm{Sym_d}$. Any homogenized elasticity tensor $\mathbf{A} \in G_\theta$ satisfies

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta \max_{\boldsymbol{\eta} \in \mathrm{Sym_d}} \left[2\boldsymbol{\sigma} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta) g^c(\boldsymbol{\eta}) \right], \quad (1.61)$$

where $g^c(\boldsymbol{\eta})$ is a nonlocal term given by

$$g^c(\boldsymbol{\eta}) = \max_{\mathsf{e} \in S^{d-1}} \left(f_2^c(\mathsf{e}) \boldsymbol{\eta} : \boldsymbol{\eta} \right), \quad (1.62)$$

where f_2^c is given by (1.50), and

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \leq \mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + (1-\theta) \min_{\boldsymbol{\eta} \in \mathrm{Sym_d}} \left[2\boldsymbol{\sigma} : \boldsymbol{\eta} + (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - \theta h^c(\boldsymbol{\eta}) \right], \quad (1.63)$$

where $h^c(\boldsymbol{\eta})$ is a nonlocal term given by

$$h^c(\boldsymbol{\eta}) = \min_{\mathsf{e} \in S^{d-1}} \left(f_1^c(\mathsf{e}) \boldsymbol{\eta} : \boldsymbol{\eta} \right),$$

where f_1^c is given by (1.58). Furthermore, these upper and lower bounds are optimal in the sense of Definition 1.6 and optimality is achieved by a finite-rank sequential laminate.

Proof. For any homogenized tensor $\mathbf{A} \in G_\theta$ there exist two sequential laminates $\mathbf{A}^+ \in L_\theta^+$ and $\mathbf{A}^- \in L_\theta^-$ such that

$$\mathbf{A}^- \leq \mathbf{A} \leq \mathbf{A}^+$$

in the sense of quadratic forms (see, for example (Allaire 2002, Theorem 2.3.11)). This implies bounds on the inverse of $\mathbf{A}$,

$$(\mathbf{A}^+)^{-1} \leq \mathbf{A}^{-1} \leq (\mathbf{A}^-)^{-1},$$

in the sense of quadratic forms, and hence

$$\mathbf{A}^{-1}\boldsymbol{\xi} : \boldsymbol{\xi} \geq (\mathbf{A}^+)^{-1}\boldsymbol{\xi} : \boldsymbol{\xi} \geq \min_{\mathbf{A}^+ \in L_\theta^+} (\mathbf{A}^+)^{-1}\boldsymbol{\xi} : \boldsymbol{\xi}, \quad \boldsymbol{\xi} \in \mathrm{Sym}_d. \tag{1.64}$$

Since L_θ^+ is closed by Lemma 1.18, it follows that (1.64) is an optimal lower bound, and it remains to evaluate the above minimum. Let us define a function

$$\varphi(\boldsymbol{\xi}) = \left((\mathbf{A}^+)^{-1} - \mathbf{A}_2^{-1}\right)\boldsymbol{\xi} : \boldsymbol{\xi}. \tag{1.65}$$

It is a convex function, and therefore, by Theorem 1.19,

$$\varphi(\boldsymbol{\xi}) = \varphi^{**}(\boldsymbol{\xi}) = \sup_{\boldsymbol{\eta} \in \mathrm{Sym}_d} [\langle \boldsymbol{\xi}, \boldsymbol{\eta} \rangle - \varphi^*(\boldsymbol{\eta})], \tag{1.66}$$

where φ^* is the Legendre transform of the function φ, i.e.

$$\varphi^*(\boldsymbol{\eta}) = \sup_{\boldsymbol{\xi} \in \mathrm{Sym}_d} [\boldsymbol{\eta} : \boldsymbol{\xi} - \left((\mathbf{A}^+)^{-1} - \mathbf{A}_2^{-1}\right)\boldsymbol{\xi} : \boldsymbol{\xi}]. \tag{1.67}$$

Note that function

$$\tilde{\varphi}(\boldsymbol{\xi}) := \boldsymbol{\eta} : \boldsymbol{\xi} - \left((\mathbf{A}^+)^{-1} - \mathbf{A}_2^{-1}\right)\boldsymbol{\xi} : \boldsymbol{\xi}$$

is concave, so the maximum in (1.67) is attained. Moreover, from the necessary condition of optimality, $\nabla\tilde{\varphi}(\boldsymbol{\xi}) = \mathbf{0}$, it follows

$$\boldsymbol{\xi} = \frac{1}{2}((\mathbf{A}^+)^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta},$$

and the Legendre transform of the function φ reads

$$\varphi^*(\boldsymbol{\eta}) = \frac{1}{4}((\mathbf{A}^+)^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta}.$$

From Lemma 1.18, there exists a probability measure ν on S^{d-1} such that

$$\theta((\mathbf{A}^+)^{-1} - \mathbf{A}_2^{-1})^{-1} = (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1} + (1-\theta)\int_{S^{d-1}} f_2^c(\mathsf{e})\, d\nu(\mathsf{e}),$$

which implies

$$\varphi^*(\boldsymbol{\eta}) = \frac{1}{4\theta}\left[(\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1} + (1-\theta)\int_{S^{d-1}} f_2^c(\mathsf{e})\, d\nu(\mathsf{e})\right]\boldsymbol{\eta} : \boldsymbol{\eta}.$$

Using (1.65) and (1.66), it follows

$$(\mathbf{A}^+)^{-1}\boldsymbol{\xi} : \boldsymbol{\xi} = \mathbf{A}_2^{-1}\boldsymbol{\xi} : \boldsymbol{\xi} + \theta \sup_{\boldsymbol{\eta}\in \mathrm{Sym}_d}\left[2\boldsymbol{\xi} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta) g_\nu(\boldsymbol{\eta})\right], \quad (1.68)$$

where

$$g_\nu(\boldsymbol{\eta}) = \left(\int_{S^{d-1}} f_2^c(\mathsf{e})\, d\nu(\mathsf{e})\right)\boldsymbol{\eta} : \boldsymbol{\eta}.$$

In order to obtain a lower bound, we minimize (1.68) with respect to probability measure $\nu \in P(S^{d-1})$. The order of minimization can be exchanged due to Theorem 1.20, because of linearity of the expression (1.68) with respect to ν and concavity with respect to $\boldsymbol{\eta}$. We obtain

$$(\mathbf{A}^+)^{-1}\boldsymbol{\xi} : \boldsymbol{\xi} = \mathbf{A}_2^{-1}\boldsymbol{\xi} : \boldsymbol{\xi} + \theta \max_{\boldsymbol{\eta}\in \mathrm{Sym}_d}\left[2\boldsymbol{\xi} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta) \max_{\nu \in P(S^{d-1})} g_\nu(\boldsymbol{\eta})\right].$$

Maximum of the function $g_\nu(\boldsymbol{\eta})$ will be attained by taking a Dirac mass concentrated on the unit vector where $f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta}$ is maximal, i.e.

$$\max_{\nu \in P(S^{d-1})} g_\nu(\boldsymbol{\eta}) = \max_{\mathsf{e}\in S^{d-1}} f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta} = g^c(\boldsymbol{\eta}).$$

This yields the lower bound (1.61). The proof for the upper bound is analogous, and we omit it. □

The nonlocal terms $g^c(\boldsymbol{\eta})$ and $h^c(\boldsymbol{\eta})$ from Proposition 1.21 can be explicitly calculated in the case of a single energy. In the lemma below, we give explicit formula for $g^c(\boldsymbol{\eta})$, while the explicit formula for $h^c(\boldsymbol{\eta})$ can be found in Allaire (2002).

Lemma 1.22 Let $\eta_1 \le \cdots \le \eta_d$ be the eigenvalues of the symmetric matrix $\boldsymbol{\eta}$. Then

$$g^c(\boldsymbol{\eta}) = \mathbf{A}_2\boldsymbol{\eta} : \boldsymbol{\eta} - \frac{1}{2\mu_2 + \lambda_2}\min\left\{(2\mu_2\eta_1 + \lambda_2 \mathrm{tr}(\boldsymbol{\eta}))^2, (2\mu_2\eta_d + \lambda_2 \mathrm{tr}(\boldsymbol{\eta}))^2\right\}.$$

Proof. By a simple calculation one can check equality

$$f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta} + f_2(\mathsf{e})\mathbf{A}_2\boldsymbol{\eta} : \mathbf{A}_2\boldsymbol{\eta} = \mathbf{A}_2\boldsymbol{\eta} : \boldsymbol{\eta},$$

and then

$$\begin{aligned} g^c(\boldsymbol{\eta}) = \max_{\mathsf{e}\in S^{d-1}} (f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta}) = \max_{\mathsf{e}\in S^{d-1}} (\mathbf{A}_2\boldsymbol{\eta} : \boldsymbol{\eta} - f_2(\mathsf{e})\mathbf{A}_2\boldsymbol{\eta} : \mathbf{A}_2\boldsymbol{\eta}) = \\ = \mathbf{A}_2\boldsymbol{\eta} : \boldsymbol{\eta} - \min_{\mathsf{e}\in S^{d-1}} f_2(\mathsf{e})\mathbf{A}_2\boldsymbol{\eta} : \mathbf{A}_2\boldsymbol{\eta}. \end{aligned}$$

Let us define a function

$$\tilde{g}(\boldsymbol{\eta}) := \min_{\mathsf{e}\in S^{d-1}} f_2(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta} = \min_{\mathsf{e}\in S^{d-1}} \left(\frac{1}{\mu_2}(|\boldsymbol{\eta}\mathsf{e}|^2 - (\boldsymbol{\eta}\mathsf{e}\cdot\mathsf{e})^2) + \frac{1}{2\mu_2+\lambda_2}(\boldsymbol{\eta}\mathsf{e}\cdot\mathsf{e})^2 \right)$$

and find the extremal value of of the above constrained minimization. By the method of Lagrange multipliers, an extremal e must satisfy the optimality condition $\nabla F = \mathbf{0}$, for $F(\mathsf{e}) = \frac{1}{\mu_2}(|\boldsymbol{\eta}\mathsf{e}|^2 - (\boldsymbol{\eta}\mathsf{e}\cdot\mathsf{e})^2) + \frac{1}{2\mu_2+\lambda_2}(\boldsymbol{\eta}\mathsf{e}\cdot\mathsf{e})^2 - l|\mathsf{e}|^2$, $l \in \mathbf{R}$ being the Lagrangian function. The optimality condition reads

$$\frac{1}{\mu_2}(\boldsymbol{\eta}^2\mathsf{e} - 2(\boldsymbol{\eta}\mathsf{e}\cdot\mathsf{e})\boldsymbol{\eta}\mathsf{e}) + \frac{2}{2\mu_2+\lambda_2}(\boldsymbol{\eta}\mathsf{e}\cdot\mathsf{e})\boldsymbol{\eta}\mathsf{e} = l\mathsf{e}. \tag{1.69}$$

This implies that $\boldsymbol{\eta}^2\mathsf{e}$ is a linear combination of vectors e and $\boldsymbol{\eta}\mathsf{e}$. Therefore, the subspace $[\mathsf{e}, \boldsymbol{\eta}\mathsf{e}]$ of $\mathbf{R}^d$ is stable under the action of $\boldsymbol{\eta} \in \mathrm{Sym}_d$. Consequently, $\boldsymbol{\eta}$ is diagonalizable on this subspace and there exists two orthogonal unit eigenvectors e_i and e_j, corresponding to the eigenvalues η_i and η_j, such that the extremal e satisfies

$$\mathsf{e} = c_i\mathsf{e}_i + c_j\mathsf{e}_j,$$

for $c_i\,, c_j \in \mathbf{R}$ and $c_i^2 + c_j^2 = 1$. By inclusion of e into (1.69) and comparing the terms with e_i and e_j, we get

$$\begin{aligned} \frac{1}{\mu_2}(c_i\eta_i^2 - 2(c_i^2\eta_i + c_j^2\eta_j)c_i\eta_i) + \frac{2}{2\mu_2+\lambda_2}(c_i^2\eta_i + c_j^2\eta_j)c_i\eta_i &= lc_i \qquad (1.70)\\ \frac{1}{\mu_2}(c_j\eta_j^2 - 2(c_i^2\eta_i + c_j^2\eta_j)c_j\eta_j) + \frac{2}{2\mu_2+\lambda_2}(c_i^2\eta_i + c_j^2\eta_j)c_j\eta_j &= lc_j. \qquad (1.71) \end{aligned}$$

If $c_i = 0$, then $\mathsf{e} = \mathsf{e}_j$, which means that e is the eigenvector corresponding to the eigenvalue η_j, and in this case

$$f_2(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta} = \frac{1}{2\mu_2+\lambda_2}\eta_j^2. \tag{1.72}$$

On the other hand, if $c_j = 0$, e is the eigenvector associated to the eigenvalue η_i and

$$f_2(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta} = \frac{1}{2\mu_2+\lambda_2}\eta_i^2. \tag{1.73}$$

If $\eta_1 = \eta_2$, we again obtain that e is the eigenvector of $\boldsymbol{\eta}$ and (1.72), i.e. (1.73) is valid. Otherwise, when $c_i \neq 0$, $c_j \neq 0$ and $\eta_i \neq \eta_j$ we simplify (1.70) by c_i and (1.71) by c_j and subtract two equalities, which yields

$$c_i^2\eta_i + c_j^2\eta_j = \frac{2\mu_2+\lambda_2}{2(\mu_2+\lambda_2)}(\eta_i + \eta_j)$$

and together with the condition $c_i^2 + c_j^2 = 1$, a linear system of two equations is obtained.

The solution of the system is

$$c_j^2 = \frac{\lambda_2\eta_i - (2\mu_2+\lambda_2)\eta_j}{2(\mu_2+\lambda_2)(\eta_i-\eta_j)}, \quad c_i^2 = \frac{(2\mu_2+\lambda_2)\eta_i - \lambda_2\eta_j}{2(\mu_2+\lambda_2)(\eta_i-\eta_j)}. \tag{1.74}$$

We can assume without loss of generality that $\eta_j > \eta_i$, which implies that $c_i^2 \geq 0$ is equivalent to $(2\mu_2+\lambda_2)\eta_i \leq \lambda_2\eta_j$, while $c_j^2 \geq 0$ is equivalent to $\lambda_2\eta_i \leq (2\mu_2+\lambda_2)\eta_j$. From these two inequalities, it follows

$$\eta_j \geq \frac{2\mu_2+\lambda_2}{2(\mu_2+\lambda_2)}(\eta_i+\eta_j) \geq \eta_i. \tag{1.75}$$

In this case

$$f_2(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta} = \frac{(\eta_i-\eta_j)^2}{4\mu_2} + \frac{(\eta_i+\eta_j)^2}{4(\mu_2+\lambda_2)} \tag{1.76}$$

and it can be easily checked that this value of $f_2(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta}$ is always greater then both values (1.72) and (1.73). To conclude, a minimum of $f_2(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta}$ is always achieved by an eigenvector of $\boldsymbol{\eta}$ and

$$\tilde{g}(\boldsymbol{\eta}) = \min_{\mathsf{e}\in S^{d-1}} f_2(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta} = \frac{1}{2\mu_2+\lambda_2}\min\{\eta_1^2, \eta_d^2\}.$$

Since the $\mathbf{A}_2$ is isotropic, the extremal vectors in the definition of $g^c(\boldsymbol{\eta})$ are the same as those for $\tilde{g}(\mathbf{A}_2\boldsymbol{\eta})$. Therefore, the extremal vector for $g^c(\boldsymbol{\eta})$ are the eigenvectors of $\mathbf{A}_2\boldsymbol{\eta} = 2\mu_2\boldsymbol{\eta} + \lambda_2\mathrm{tr}(\boldsymbol{\eta})\mathbf{I}_2$, which are also eigenvectors of $\boldsymbol{\eta}$ and

$$g^c(\boldsymbol{\eta}) = \mathbf{A}_2\boldsymbol{\eta} : \boldsymbol{\eta} - \frac{1}{2\mu_2+\lambda_2}\min\left\{(2\mu_2\eta_1 + \lambda_2\mathrm{tr}(\boldsymbol{\eta}))^2, (2\mu_2\eta_d + \lambda_2\mathrm{tr}(\boldsymbol{\eta}))^2\right\}.$$

□

As in the conductivity case, equality in Hashin-Shtrikman bounds can be achieved with sequential laminates. To show that, we need to employ a theory of subgradients. Let us briefly describe it, while for more information we refer to Schirotzek (2007), Rockafellar (1972) and Dacorogna (2008).

Definition 1.8 Let X be a Banach space and $f : X \to \overline{\mathbf{R}}$ a convex function such that $f(\mathbf{x}) > -\infty$, $\mathbf{x} \in X$. The subdifferential of the function f at $\overline{\mathbf{x}} \in X$ is defined as

$$\partial f(\overline{\mathbf{x}}) := \{\xi \in X' : {}_{X'}\langle \xi, \mathbf{x} - \overline{\mathbf{x}}\rangle_X \leq f(\mathbf{x}) - f(\overline{\mathbf{x}}),\, \mathbf{x} \in X\}.$$

Each $\xi \in \partial f(\overline{\mathbf{x}})$ is called subgradient of f at $\overline{\mathbf{x}}$.

Note that if the function f is continuosly differentiable, then $\partial f(\overline{\mathbf{x}})$ reduces to the singleton set $\{\nabla f(\overline{\mathbf{x}})\}$. The main pourpose of subdifferential is to detect minimum points of the function.

Proposition 1.23 If $f : X \to \overline{\mathbf{R}}$ is convex, then f attains a minimum in $\overline{\mathbf{x}} \in X$ if and only if $0 \in \partial f(\overline{\mathbf{x}})$.

$\square$

Theorem 1.24 Let S be a compact Hausdorff space. For any $s \in S$ let $f_s : X \to \mathbf{R}$ be convex on X and continuous at $\overline{\mathbf{x}} \in X$. Assume further that there exists a neighborhood U of $\overline{\mathbf{x}}$ such that for every $\mathbf{y} \in U$ the functional $s \mapsto f_s(\mathbf{y})$ is upper semicontinuous on S. Then the functional $f : X \to \mathbf{R}$ defined by

$$f(\mathbf{x}) = \max_{s \in S} f_s(\mathbf{x}), \quad x \in X$$

satisfies

$$\partial f(\overline{x}) = \overline{\mathrm{co}}^* \left(\bigcup_{s \in S(\overline{\mathbf{x}})} \partial f_s(\overline{\mathbf{x}}) \right),$$

where $S(\overline{\mathbf{x}}) = \{s \in S : f_s(\overline{\mathbf{x}}) = f(\overline{\mathbf{x}})\}$, while $\overline{\mathrm{co}}^* M$ is the weak-$*$ closure of the convex hull of $M \subseteq X'$.

$\square$

Let us also recall one of the most important characterization of convex hulls, i.e. Carathéodory's theorem.

Theorem 1.25 Let $E \subseteq \mathbf{R}^n$. Then

$$\mathrm{co}\, E = \left\{ \sum_{i=1}^{n+1} m_i \mathbf{x}_i : \mathbf{x}_i \in E, m_i \geq 0, \sum_{i=1}^{n+1} m_i = 1 \right\}.$$

$\square$

Proposition 1.26 Optimality in Hashin-Shtrikman bounds (1.61) and (1.63) can be achieved by a rank-d sequential laminate with the lamination directions that are extremal in the definition of the nonlocal terms $g^c(\boldsymbol{\eta})$ and $h^c(\boldsymbol{\eta})$. In particular, the optimal rank-d sequential laminate for the lower bound is aligned with the eigendirections of $\boldsymbol{\sigma}$.

Proof. The equality in Hashin-Shtrikman bound is checked by inspecting the condition of optimality with respect to $\boldsymbol{\eta}$. Let us first consider the lower bound (1.61) and rewrite it as

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta \max_{\boldsymbol{\eta} \in \mathrm{Sym_d}} \phi_{\boldsymbol{\sigma}}(\boldsymbol{\eta}),$$

where $\phi_{\boldsymbol{\sigma}}(\boldsymbol{\eta}) = 2\boldsymbol{\sigma} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta)g^c(\boldsymbol{\eta})$. Since $f_2^c(\mathsf{e})$ is positive semidefinite, the mapping $\boldsymbol{\eta} \mapsto f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta}$ is convex, and therefore

$$g^c(\boldsymbol{\eta}) = \max_{\mathsf{e} \in S^{d-1}} (f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta})$$

is a convex function implying that $\phi_{\boldsymbol{\sigma}}$ is concave in $\boldsymbol{\eta}$ (and thus $-\phi_{\boldsymbol{\sigma}}$ is convex). Since $\phi_{\boldsymbol{\sigma}}$ is not a smooth function in general, the necessary condition of optimality reads

$$\mathbf{0} \in \partial\phi_{\boldsymbol{\sigma}}(\boldsymbol{\eta}^*), \tag{1.77}$$

where $\partial\phi_{\boldsymbol{\sigma}}(\boldsymbol{\eta}^*)$ is the subdifferential of the function $\phi_{\boldsymbol{\sigma}}$ at the optimal point $\boldsymbol{\eta}^*$. To calculate subdifferential of the function $\phi_{\boldsymbol{\sigma}}$, we need to calculate subdifferential of the function g^c. Notice that g^c is defined as a maximum of a family of functions parametrized by e on the compact set, which satisfy assumptions of Theorem 1.24. Moreover, since $\boldsymbol{\eta} \in \mathrm{Sym}_d$, which is of finite dimension, Theorem 1.24 implies

$$\partial g^c(\boldsymbol{\eta}) = \mathrm{co}\left(\{2f_2^c(\mathsf{e})\boldsymbol{\eta} : \mathsf{e} \in S(\boldsymbol{\eta})\}\right), \tag{1.78}$$

where $S(\boldsymbol{\eta}) \subseteq S^{d-1}$ is the set of all extremal e in the definition of $g(\boldsymbol{\eta})$. Using Theorem 1.25, (1.78) becomes

$$\partial g^c(\boldsymbol{\eta}) = \left\{2\sum_{i=1}^{p} m_i f_2^c(\mathsf{e}_i)\boldsymbol{\eta} : \mathsf{e}_i \in S(\boldsymbol{\eta})\right\},$$

for $m_i \geq 0$, $\sum_{i=1}^{p} m_i = 1$ and $p = \frac{d(d+1)}{2} + 1$. Hence, the optimality condition (1.77) implies

$$\boldsymbol{\sigma} = (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta}^* + (1-\theta)\sum_{i=1}^{p} m_i f_2^c(\mathsf{e}_i)\boldsymbol{\eta}^*. \tag{1.79}$$

From Lemma 1.22 these extremal vectors e_i correspond to the same eigenvalue, η_1^* or η_d^*, of $\boldsymbol{\eta}^*$ and (1.79) becomes

$$\begin{aligned}\boldsymbol{\sigma} =& (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta}^* + (1-\theta)\left(\mathbf{A}_2\boldsymbol{\eta}^* - \frac{\lambda_2(2\mu_2\eta_q^* + \lambda_2 \mathrm{tr}\boldsymbol{\eta}^*)}{2\mu_2+\lambda_2}\mathbf{I}\right) - \\ & -(1-\theta)\frac{2\mu_2(2\mu_2\eta_q^* + \lambda_2\mathrm{tr}\boldsymbol{\eta}^*)}{2\mu_2+\lambda_2}\sum_{i=1}^{p} m_i \mathsf{e}_i \otimes \mathsf{e}_i,\end{aligned}$$

with $q = 1$ or $q = d$. The symmetric matrix $P = \sum_{i=1}^{p} m_i \mathsf{e}_i \otimes \mathsf{e}_i$ is positive semidefinite and of trace 1, therefore it can be diagonalized as

$$P = \sum_{i=1}^{d} \tilde{m}_i \tilde{\mathsf{e}}_i \otimes \tilde{\mathsf{e}}_i,$$

with unit vectors $\tilde{\mathsf{e}}_i$ and eigenvalues $\tilde{m}_i$, such that $\tilde{m}_i \geq 0$ and $\sum_{i=1}^{d} \tilde{m}_i = 1$. Since e_i belongs to the same eigenspace for $\boldsymbol{\eta}^*$, so does the $\tilde{\mathsf{e}}_i$, and they are also extremal for $g^c(\boldsymbol{\eta}^*)$. Thus, we can consider $p \leq d$ in the optimality condition (1.79). In order to achieve equality in the lower Hashin-Shtrikman bound (1.61), consider the sequential laminate with core $\mathbf{A}_1$ and matrix $\mathbf{A}_2$, parameters $\tilde{m}_i$ and lamination directions $\tilde{\mathsf{e}}_i$:

$$\theta(\mathbf{A}^{-1} - \mathbf{A}_2^{-1})^{-1} = (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1} + (1-\theta)\sum_{i=1}^{d} \tilde{m}_i f_2^c(\mathsf{e}_i).$$

By multiplying the above by $\boldsymbol{\eta}^*$ and using the necessary condition of optimality (1.79), we get

$$\theta\boldsymbol{\eta}^* = \mathbf{A}^{-1}\boldsymbol{\sigma} - \mathbf{A}_2^{-1}\boldsymbol{\sigma},$$

which, upon taking the inner product with $\boldsymbol{\sigma}$, implies

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} = \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta\boldsymbol{\eta}^* : \boldsymbol{\sigma}. \tag{1.80}$$

Furthermore, by taking the inner product with $\boldsymbol{\eta}^*$ in the necessary condition of optimality (1.79), we get

$$\boldsymbol{\sigma} : \boldsymbol{\eta}^* = (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta}^* : \boldsymbol{\eta}^* + (1-\theta)g^c(\boldsymbol{\eta}^*),$$

which implies

$$\boldsymbol{\eta}^* : \boldsymbol{\sigma} = \phi_{\boldsymbol{\sigma}}(\boldsymbol{\eta}^*). \tag{1.81}$$

Equation (1.81) together with (1.80) proves the assertion of the proposition for the lower Hashin-Shtrikman bound. For the proof of the upper bound we refer to Allaire (2002). □

We already commented the lack of knowledge of the G-closure set in elasticity setting. This is major difficulty in application of the homogenization method in optimal design problems. However, there exists a restricted class of optimal design problems where the homogenization theory has found its applications, due to sequential laminates which appear as optimal in this class of problems.

Chapter 2

Optimal design problems in conductivity

As was mentioned in the introduction, in optimal design problems we are seeking for a distribution of given materials within the domain Ω which minimizes an integral functional of form

$$I(\boldsymbol{\chi}) = \int_\Omega F(\mathbf{x}, u(\mathbf{x}), \boldsymbol{\chi}(\mathbf{x}))\, d\mathbf{x}, \tag{2.1}$$

where $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_k) \in \mathrm{L}^\infty(\Omega; K)$, $K = \{\boldsymbol{\kappa} \in \{0,1\}^k : \sum_{j=1}^k \kappa_j = 1\}$ describes a rearrangement of materials that constitute the domain, while the function u is a solution of a partial differential equation modelling the involved physics. The functional I is called an *objective function*, and the corresponding partial differential equation is called the *state equation*. In this section the state equation is the stationary diffusion equation

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u) = f & \text{in } \Omega, \\ u \in \mathrm{H}_0^1(\Omega), \end{cases}$$

modelling thermal (or electrical) conductivity. Since the temperature u depends on external heat density f, it follows that an optimal distribution of materials in the domain also depends on the right-hand side of the associated state equation. In applications, it is often necessary to make a structure which will work properly in different regimes. Therefore, it is more convenient to consider optimal design problems with several state equations, called *multiple state* optimal design problems and we shall deal with them in the sequel. Additionally, we shall restrict ourselves to mixtures of two isotropic materials.

2.1 Multiple state optimal design problems

Let $\Omega \subseteq \mathbf{R}^d$ be an open and bounded domain, filled with two isotropic materials with conductivities $0 < \alpha < \beta$. If we denote by $\chi \in \mathrm{L}^\infty(\Omega; \{0,1\})$ a characteristic function of the part of the domain occupied by the first material, the one with conductivity α, then the conductivity of the mixture is given by

$$\mathbf{A} = \chi\alpha\mathbf{I} + (1-\chi)\beta\mathbf{I}.$$

Moreover, let us assume that the volume of the first material is prescribed: $\int_\Omega \chi(\mathbf{x})\,d\mathbf{x} = q_\alpha$, where $0 < q_\alpha < |\Omega|$ is given. A multiple state optimal design problem consists of minimizing the functional

$$I(\chi) = \int_\Omega \left(\chi(\mathbf{x})g_\alpha(\mathbf{x},\mathsf{u}) + (1-\chi(\mathbf{x}))g_\beta(\mathbf{x},\mathsf{u})\right)\,d\mathbf{x}\,, \tag{2.2}$$

where functions g_α and g_β are given functions, over the set of all measurable characteristic functions on Ω satisfying the volume constraint. Here, a function $\mathsf{u} = (u_1,\ldots,u_m)$ is a vector function made of the solutions of equations

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u_i) = f_i & \text{in } \Omega, \\ u_i \in \mathrm{H}^1_0(\Omega)\,, \end{cases} \tag{2.3}$$

for some given $f_i \in \mathrm{H}^{-1}(\Omega)$, $i = 1,\ldots,m$.

Remark 2.1. Functional (2.2) corresponds to the functional (2.1), but it is written in a more convenient form. Since χ takes values in $\{0,1\}$, we define $g_\alpha(\mathbf{x},\mathsf{u}) = F(\mathbf{x},\mathsf{u},1)$ and $g_\beta(\mathbf{x},\mathsf{u}) = F(\mathbf{x},\mathsf{u},0)$ and write (2.1) in form (2.2).

The volume constraint of the first material in (2.2) is handled by introducing a Lagrange multiplier l, leading to an unconstrained minimization problem

$$\begin{cases} J(\chi) = \int_\Omega \left(\chi(\mathbf{x})g_\alpha(\mathbf{x},\mathsf{u}) + (1-\chi(\mathbf{x}))g_\beta(\mathbf{x},\mathsf{u})\right)\,d\mathbf{x} + l\int_\Omega \chi(\mathbf{x})\,d\mathbf{x} \longrightarrow \min \\ \chi \in L^\infty(\Omega;\{0,1\}). \end{cases} \tag{2.4}$$

The classical method for obtaining the existence of optimal design in (2.4) is the direct method of the calculus of variations. Let us describe how the method should be used, under the assumption that functional J is bounded below: consider a minimizing sequence (χ^n) in the set of admissible designs $L^\infty(\Omega;\{0,1\})$ such that

$$\lim_{n\to\infty} J(\chi^n) = \inf_{\chi\in\mathrm{L}^\infty(\Omega;\{0,1\})} J(\chi).$$

If one can prove that the sequence χ^n converges to a limit χ^∞ (at least on a subsequence) in a suitable topology, and that J is sequentially lower semicontinuous, then

$$\inf_{\chi\in\mathrm{L}^\infty(\Omega;\{0,1\})} J(\chi) = \lim_{n\to\infty} J(\chi^n) \geq J(\chi^\infty) \geq \inf_{\chi\in\mathrm{L}^\infty(\Omega;\{0,1\})} J(\chi),$$

which implies that χ^∞ is a minimizer of the functional J. Typical topologies considered on the space $L^\infty(\Omega;\{0,1\})$ are strong (or pointwise) convergence and the weak-$*$ convergence. Unfortunately, $L^\infty(\Omega;\{0,1\})$ is not sequentially compact space with either of these topologies, and therefore the direct method of calculus of variations cannot be applied. There exist various counterexamples for existence of classical solutions (Murat (1971), Murat & Tartar (1985), Lurie et al. (1982), Allaire (2002)). In order to obtain existence of a solution, problem (2.4) must be relaxed in some sense.

2.1.1 Relaxation by the homogenization method

For the relaxation of the original problem (2.4), we need to find the closure of the space of admissible designs and extend the objective function to this closure. Taking weak-$*$ closure of $\mathrm{L}^\infty(\Omega; \{0,1\})$, i.e. $\mathrm{L}^\infty(\Omega; [0,1])$ as a space of generalized designs would ensure compactness, but mapping $\chi \mapsto \mathsf{u}$ (through the state equation) is not continuous in this space. Following Murat and Tartar's homogenization method (see Chapter 1), a proper generalized (composite) design is a couple $(\theta, \mathbf{A})$, made of local proportion θ and homogenized conductivity matrix $\mathbf{A}$, i.e. composite material defined in Definition 1.2. From Theorem 1.13, the set of all generalized designs is

$$\mathcal{A} = \{(\theta, \mathbf{A}) \in L^\infty(\Omega; [0,1] \times \mathcal{M}_{\alpha,\beta}) : \mathbf{A}(\mathbf{x}) \in \mathcal{K}(\theta(\mathbf{x})) \text{ a.e. } \mathbf{x} \in \Omega\}, \tag{2.5}$$

where the set $\mathcal{K}(\theta)$ of all possible homogenized conductivities which can be obtained with the prescribed local fraction θ consists of all symmetric matrices with eigenvalues $\lambda_1, \dots, \lambda_d$ satisfying the inequalities

$$\lambda_\theta^- \leq \lambda_j \leq \lambda_\theta^+ , \quad j = 1, \dots, d, \tag{2.6}$$

$$\sum_{j=1}^d \frac{1}{\lambda_j - \alpha} \leq \frac{1}{\lambda_\theta^- - \alpha} + \frac{d-1}{\lambda_\theta^+ - \alpha}, \tag{2.7}$$

$$\sum_{j=1}^d \frac{1}{\beta - \lambda_j} \leq \frac{1}{\beta - \lambda_\theta^-} + \frac{d-1}{\beta - \lambda_\theta^+}, \tag{2.8}$$

where $\lambda_\theta^- = \left(\frac{\theta}{\alpha} + \frac{1-\theta}{\beta}\right)^{-1}$ and $\lambda_\theta^+ = \theta\alpha + (1-\theta)\beta$.

Since $L^\infty(\Omega; [0,1] \times \mathcal{M}_{\alpha,\beta})$ is compact with respect to the weak-$*$ topology for θ and H-topology for $\mathbf{A}$, the set $\mathcal{A}$ is compact, and the relaxation of problem (2.4) reads

$$\begin{cases} J(\theta, \mathbf{A}) = \int_\Omega \left(\theta(\mathbf{x}) g_\alpha(\mathbf{x}, \mathsf{u}) + (1-\theta(\mathbf{x})) g_\beta(\mathbf{x}, \mathsf{u})\right) \, d\mathbf{x} + l \int_\Omega \theta(\mathbf{x}) \, d\mathbf{x} \longrightarrow \min \\ (\theta, \mathbf{A}) \in \mathcal{A}, \end{cases} \tag{2.9}$$

where u is the vector of solutions of state equations (2.3). Moreover, continuity of the functional J from (2.9) with respect to the product topology listed above can be obtained with the following assumptions on functions g_α and g_β:

$$\begin{cases} \mathbf{x} \mapsto g_{\alpha,\beta}(\mathbf{x}, \mathsf{v}) \text{ are measurable for any } \mathsf{v} \in \mathbf{R}^d \\ \mathsf{v} \mapsto g_{\alpha,\beta}(\mathbf{x}, \mathsf{v}) \text{ are continuous for a.e. } \mathbf{x} \in \Omega \\ |g_{\alpha,\beta}(\mathbf{x}, \mathsf{v})| \le a|\mathsf{v}|^s + b(\mathbf{x}), \text{for some } a > 0,\ b \in \mathrm{L}^1(\Omega), \text{ and } 1 \le s < \frac{2d}{d-2}. \end{cases} \tag{2.10}$$

The first two assumptions on $g_{\alpha,\beta}$ in (2.10) correspond to Carathéodory functions. In dimensions $d = 1$ or $d = 2$, the exponent s in the third assumption has to be understood in the sense that $1 \le s < \infty$. These assumptions ensure that the functions $\mathsf{u}(\cdot) \mapsto g_\alpha(\cdot, \mathsf{u}(\cdot))$ and $\mathsf{u}(\cdot) \mapsto g_\beta(\cdot, \mathsf{u}(\cdot))$ are continuous from $\mathrm{H}^1_0(\Omega)$ with weak topology into $\mathrm{L}^1(\Omega)$ with strong topology. To see this, we use the following lemma (Dudley & Norvaiša (2011)).

Lemma 2.1 Let Ω be a measurable set in $\mathbf{R}^d$ and $f : \Omega \times \mathbf{R}^d \to \mathbf{R}$ a Carathéodory function for which there exists $a \ge 0$ and $b \in \mathrm{L}^q(\Omega)$ such that

$$f(\mathbf{x}, \mathsf{u}) \le a|\mathsf{v}|^{\frac{p}{q}} + b(\mathbf{x}), \quad \text{a.e.}\, \mathbf{x} \in \Omega,\ \mathsf{v} \in \mathbf{R}^d.$$

Then Nemytckii operator N_f, which maps function u to function $f(\cdot, \mathsf{u})$ is continuous from $\mathrm{L}^p(\Omega)$ into $\mathrm{L}^q(\Omega)$, $1 \le p, q < \infty$.

□

For the functions g_α and g_β, Nemytskii operators N_{g_α} and N_{g_β} map continuously from $\mathrm{L}^s(\Omega)$ into $\mathrm{L}^1(\Omega)$, and since $\mathrm{H}^1_0(\Omega)$ is compactly imbedded into $\mathrm{L}^s(\Omega)$ for $1 \le s < \frac{2d}{d-2}$, it follows that N_{g_α} and N_{g_β} are continuous from $\mathrm{H}^1_0(\Omega)$ with weak topology into $\mathrm{L}^1(\Omega)$ with strong topology. As a cosequence, the following lemma is valid.

Lemma 2.2 Under the assumptions (2.10), the functional J from (2.9) is continuous on $\mathcal{A}$ with weak-$*$ topology for θ and H-topology for $\mathbf{A}$.

Proof. Since the topology defined on the domain of J is metrizable, it is sufficient to prove sequential continuity of the functional J. Let $(\theta^n, \mathbf{A}^n)$ be a sequence in $\mathcal{A}$ such that

$$\begin{aligned} \theta^n &\xrightharpoonup{*} \theta \quad \text{in } \mathrm{L}^\infty(\Omega; [0,1]) \\ \mathbf{A}^n &\xrightharpoonup{H} \mathbf{A} \quad \text{in } \mathrm{L}^\infty(\Omega, \mathcal{M}_{\alpha,\beta}). \end{aligned}$$

From the definition of H-convergence, for the corresponding sequence $\mathsf{u}^n = (u^n_1, \dots, u^n_m)$ of vectors of solutions of (2.3) with $\mathbf{A}^n$ instead of $\mathbf{A}$, the following is valid

$$\mathsf{u}^n \rightharpoonup \mathsf{u} \quad \text{in } \mathrm{H}^1_0(\Omega; \mathbf{R}^{\mathrm{d}}),$$

and u is the vector of solutions of state equations (2.3). Moreover, since the embedding of $\mathrm{H}^1_0(\Omega)$ into $\mathrm{L}^2(\Omega)$ is compact, the sequence (u^n) converges strongly to u in $\mathrm{L}^2(\Omega; \mathbf{R}^d)$.

By Lemma 2.1

$$g_k(\cdot, \mathsf{u}^n(\cdot)) \longrightarrow g_k(\cdot, \mathsf{u}(\cdot)) \quad \text{in } \mathrm{L}^1(\Omega), \quad k \in \{\alpha, \beta\}.$$

Since the dual product of weakly and strongly converging sequences converges to the dual product of the corresponding limits, it follows

$$\int_\Omega \theta^n(\mathbf{x}) g_k(\mathbf{x}, \mathsf{u}^n(\mathbf{x}))\, d\mathbf{x} \longrightarrow \int_\Omega \theta(\mathbf{x}) g_k(\mathbf{x}, \mathsf{u}(\mathbf{x}))\, d\mathbf{x}, \quad k \in \{\alpha, \beta\},$$

and therefore,

$$J(\theta^n, \mathbf{A}^n) \longrightarrow J(\theta, \mathbf{A}),$$

which concludes the proof. □

This relaxation does not modify the problem substantially, meaning that there is still a connection between classical and generalized solutions, which is stated in the following theorem (Murat & Tartar (1985)).

Theorem 2.3 The minimization problem (2.9) is a proper relaxation of the original problem (2.4) in the sense that

1. there exists at least one minimizer of J on $\mathcal{A}$,

2. up to a subsequence, every minimizing sequence of classical designs χ^n for J weak-$*$ converges in $\mathrm{L}^\infty(\Omega; [0,1])$ to a density function θ, and the associated conductivitiy $\mathbf{A}^n = \chi^n \alpha \mathbf{I} + (1 - \chi^n)\beta \mathbf{I}$ H-conveges to a homogenized conductivity tensor $\mathbf{A}$ such that $(\theta, \mathbf{A})$ is a minimizer of J on $\mathcal{A}$,

3. conversely, every minimizer $(\theta, \mathbf{A})$ of J on $\mathcal{A}$ is attained by a minimizing sequence, for J, of classical designs χ^n, namely θ is the weak-$*$ limit of χ^n in $\mathrm{L}^\infty(\Omega; [0,1])$ and $\mathbf{A}$ is the H-limit of $\mathbf{A}^n = \chi^n \alpha \mathbf{I} + (1 - \chi^n)\beta \mathbf{I}$.

Proof. Let (χ^n) be a minimizing sequence of characteristic functions for the problem (2.4). Since it is bounded in $\mathrm{L}^\infty(\Omega; [0,1])$, there exists a subsequence, denoted the same, such that

$$\chi^n \xrightharpoonup{*} \theta \quad \text{in } \mathrm{L}^\infty(\Omega; [0,1]).$$

Moreover, from the compactness of H-convergence stated in Theorem 1.1, for the sequence $\mathbf{A}^n = \chi^n \alpha \mathbf{I} + (1 - \chi^n)\beta \mathbf{I}$ there exists a subsequence, denoted the same, such that $\mathbf{A}^n \xrightharpoonup{H} \mathbf{A}$ and then Lemma 2.2 implies

$$\lim_{n \to \infty} J(\chi^n) = J(\theta, \mathbf{A}).$$

Since χ^n is a minimizing sequence for (2.4), it follows

$$J(\theta, \mathbf{A}) = \inf_{\chi \in \mathrm{L}^\infty(\Omega; \{0,1\})} J(\chi).$$

Furthermore, any generalized composite design $(\tilde{\theta}, \tilde{\mathbf{A}}) \in \mathcal{A}$ is attained as the limit of a sequence of classical designs $\tilde{\chi}^n \in \mathrm{L}^\infty(\Omega; \{0,1\})$, in the sense that $\tilde{\theta}$ is the weak-$*$ limit of $\tilde{\chi}^n$ in $\mathrm{L}^\infty(\Omega; \{0,1\})$, and $\tilde{\mathbf{A}}$ is the H-limit of the sequence $\mathbf{A}^n = \tilde{\chi}^n \alpha \mathbf{I} + (1-\tilde{\chi}^n)\beta\mathbf{I}$. Thus, for any $(\tilde{\theta}, \tilde{\mathbf{A}}) \in \mathcal{A}$, there exists a sequence $\tilde{\chi}^n \in \mathrm{L}^\infty(\Omega; [0,1])$ such that

$$J(\tilde{\theta}, \tilde{\mathbf{A}}) = \lim_{n\to\infty} J(\tilde{\chi}^n) \geq \inf_{\chi \in \mathrm{L}^\infty(\Omega;[0,1])} J(\chi) = J(\theta, \mathbf{A}).$$

We conclude that $(\theta, \mathbf{A})$ is a minimizer of J on $\mathcal{A}$, which proves the first two statements of the theorem. For the last statement, if $(\theta, \mathbf{A})$ is a minimizer for J on $\mathcal{A}$, then there exists a sequence $\chi^n \in \mathrm{L}^\infty(\Omega; \{0,1\})$ which weak-$*$ converges to θ in $\mathrm{L}^\infty(\Omega; [0,1])$, while the sequence $\chi^n\alpha\mathbf{I} + (1-\chi^n)\beta\mathbf{I}$ H-converges to $\mathbf{A}$. If $\tilde{\chi}^n$ is a minimizing sequence for J in $\mathrm{L}^\infty(\Omega; [0,1])$, then from the second assertion of the theorem it follows

$$\inf_{\chi \in \mathrm{L}^\infty(\Omega;[0,1])} J(\chi) = \lim_{n\to\infty} J(\tilde{\chi}^n) = J(\theta, \mathbf{A}) = \lim_{n\to\infty} J(\chi^n).$$

Hence, the sequence χ^n is a minimizing sequence for J in $\mathrm{L}^\infty(\Omega; [0,1])$ and the proof is complete. □

This relaxation process enables us to prove the existence of relaxed optimal design, but in general, it is not unique. For example, consider a single state energy minimization problem on a ball $B(\mathbf{0}, R)$, with the right-hand side equal to a characteristic function of an annulus $B(\mathbf{0}; r, R)$, for $r \in [0, R]$. If we have enough material α to fill the annulus, then the rest of the material α can be distributed within the smaller ball $B(\mathbf{0}, R)$ in an arbitrary way and any such design will be optimal for the considered problem (Burazin (2018)). Additionally, although the set $\mathcal{K}(\theta)$ is convex for any $\theta \in [0,1]$, the set $\mathcal{A}$ is not convex, due to the upper bound (2.8) (Vrdoljak (2016)). This implies that the necessary conditions of optimality for the relaxed formulation are not sufficient in general, and local minima need not be global minima.

2.1.2 Necessary conditions of optimality

We shall derive the necessary conditions of optimality in terms of the Gâteaux differential (we use an analogous technique to that presented in Murat & Tartar (1985), Tartar (1995, 2000) and Allaire (2002)). Let us denote by $(\theta^*, \mathbf{A}^*)$ a local minimum of the relaxed problem (2.9). A point $(\delta\theta, \delta\mathbf{A})$ is said to be an admissible direction (variation) in $(\theta^*, \mathbf{A}^*)$ if for small $\varepsilon > 0$

$$(\theta, \mathbf{A}) + \varepsilon(\delta\theta, \delta\mathbf{A}) + o(\varepsilon) \in \mathcal{A}, \quad \lim_{\varepsilon \searrow 0} \frac{\|o(\varepsilon)\|}{\varepsilon} = 0.$$

Furthermore, a path $\varepsilon \mapsto (\theta^\varepsilon, \mathbf{A}^\varepsilon)$ is called admissible in $\mathcal{A}$ if $(\theta^\varepsilon, \mathbf{A}^\varepsilon)$ is admissible, for every $\varepsilon > 0$. If we consider an admissible smooth path in $\mathcal{A}$ given by

$$\varepsilon \mapsto (\theta^\varepsilon, \mathbf{A}^\varepsilon) = (\theta^*, \mathbf{A}^*) + \varepsilon(\delta\theta, \delta\mathbf{A}) + o(\varepsilon), \tag{2.11}$$

which passes through $(\theta^*, \mathbf{A}^*)$ for $\varepsilon = 0$, then an admissible direction reads

$$(\delta\theta, \delta\mathbf{A}) = \frac{d}{d\varepsilon}(\theta^\varepsilon, \mathbf{A}^\varepsilon)\big|_{\varepsilon=0^+}.$$

To ensure Gâteaux differentiability of the objective function $J(\theta, \mathbf{A})$, we assume that for any $i = 1, \dots, m$ the partial derivatives $\frac{\partial g_\alpha}{\partial u_i}$ and $\frac{\partial g_\beta}{\partial u_i}$ are Carathéodory functions satisfying the growth condition

$$\left|\frac{\partial g_{\alpha,\beta}}{\partial u_i}\right| \le a'|\mathsf{u}|^{s-1} + b'(\mathbf{x}) \quad \text{a.e. } \mathbf{x} \in \Omega,$$

where $a' \ge 0$ and $b' \in \mathrm{L}^q(\Omega)$, $q \ge \frac{2d}{d+2}$, $1 \le s \le \frac{2d}{d-2}$.

Theorem 2.4 Let $(\theta^*, \mathbf{A}^*)$ be the minimizer of objective functional $J(\theta, \mathbf{A})$ in (2.9) with states u_i^* and corresponding adjoint states p_i^*, $i = 1, \dots m$, defined as the unique solutions of the adjoint boundary value problems

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^*\nabla p_i^*) = \theta^* \dfrac{\partial g_\alpha}{\partial u_i}(\cdot, \mathsf{u}^*) + (1-\theta^*)\dfrac{\partial g_\beta}{\partial u_i}(\cdot, \mathsf{u}^*) \quad \text{in } \Omega \\ p_i^* \in \mathrm{H}_0^1(\Omega)\,. \end{cases} \tag{2.12}$$

The necessary condition of optimality for the functional J reads

$$\int_\Omega (g_\alpha(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) - g_\beta(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) + l)\,\delta\theta(\mathbf{x})\,d\mathbf{x} - \int_\Omega \sum_{i=1}^m \delta\mathbf{A}(\mathbf{x})\nabla u_i^*(\mathbf{x})\cdot\nabla p_i^*(\mathbf{x})\,d\mathbf{x} \ge 0, \tag{2.13}$$

for any admissible variation $(\delta\theta, \delta\mathbf{A}) = \frac{d}{d\varepsilon}(\theta^\varepsilon, \mathbf{A}^\varepsilon)\big|_{\varepsilon=0^+}$.

Proof. Let $\varepsilon \mapsto (\theta^\varepsilon, \mathbf{A}^\varepsilon)$ be the smooth admissible path given by (2.11), and $\mathsf{u}^\varepsilon = (u_1^\varepsilon, u_2^\varepsilon, \dots, u_m^\varepsilon)$ the vector of solutions of (2.3) with the design parameters $(\theta^\varepsilon, \mathbf{A}^\varepsilon)$. Since $(\theta^*, \mathbf{A}^*)$ is the minimum point of the functional J, it follows

$$\frac{d}{d\varepsilon} J(\theta^\varepsilon, \mathbf{A}^\varepsilon)\big|_{\varepsilon=0^+} \ge 0. \tag{2.14}$$

Since state functions appear in the definition of the functional J, for the calculation of the left-hand side in (2.14), we also need to calculate $\frac{du_i^\varepsilon}{d\varepsilon}\big|_{\varepsilon=0^+}$. If we define $\delta u_1, \dots, \delta u_m$ as solutions of equations

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^*\nabla\delta u_i) = \mathsf{div}\,(\delta\mathbf{A}\nabla u_i^*) \quad \text{in } \Omega \\ \delta u_i \in \mathrm{H}_0^1(\Omega) \end{cases} \tag{2.15}$$

then $\frac{du_i^\varepsilon}{d\varepsilon}\big|_{\varepsilon=0^+} = \delta u_i$, $i = 1, \dots, m$. Indeed, multiplying the above equation by ε, and recalling $-\mathsf{div}\,(\mathbf{A}^*\nabla u_i^*) = f_i$, we get

$$-\mathsf{div}\,(\mathbf{A}^* + \varepsilon\delta\mathbf{A})(\nabla u_i^* + \varepsilon\nabla\delta u_i) = f_i - \varepsilon^2\mathsf{div}\,(\delta\mathbf{A}\nabla\delta u_i),$$

i.e.

$$-\mathsf{div}\,\mathbf{A}^\varepsilon(\nabla u_i^* + \varepsilon\nabla\delta u_i) = f_i + o(\varepsilon),$$

where $\lim_{\varepsilon\searrow 0}\frac{\|o(\varepsilon)\|_{\mathrm{H}^{-1}(\Omega)}}{\varepsilon} = 0$. Subtracting the state equation $-\mathsf{div}\,(\mathbf{A}^\varepsilon u_i^\varepsilon) = f_i$ from the above equation, it follows

$$-\mathsf{div}\,(\mathbf{A}^\varepsilon(\nabla u_i^\varepsilon - \nabla u_i^* - \varepsilon\nabla\delta u_i)) = o(\varepsilon).$$

From the a priori estimate (1.2), the inequality

$$\|u_i^\varepsilon - u_i^* - \varepsilon\delta u_i\|_{\mathrm{H}_0^1(\Omega)} \le \frac{1}{\alpha}\|o(\varepsilon)\|_{\mathrm{H}^{-1}(\Omega)}$$

follows, which implies $\frac{du_i^\varepsilon}{d\varepsilon}\big|_{\varepsilon=0^+} = \delta u_i$, $i = 1, \dots, m$. In particular, mapping $\varepsilon \mapsto \mathsf{u}^\varepsilon$ from $[0, 1]$ to $\mathrm{H}_0^1(\Omega; \mathbf{R}^d)$ with strong topology is continuous at zero. We now calculate

$$\begin{aligned}\frac{J(\theta^\varepsilon, \mathbf{A}^\varepsilon) - J(\theta^*, \mathbf{A}^*)}{\varepsilon} &= \int_\Omega \frac{\theta^\varepsilon - \theta^*}{\varepsilon}\left[g_\alpha(\mathbf{x}, \mathsf{u}^\varepsilon) - g_\beta(\mathbf{x}, \mathsf{u}^\varepsilon) + l\right]\,d\mathbf{x}+\\ &+ \int_\Omega \frac{1}{\varepsilon}\theta^*(g_\alpha(\mathbf{x}, \mathsf{u}^\varepsilon) - g_\alpha(\mathbf{x}, \mathsf{u}^*))\,d\mathbf{x}+\\ &+ \int_\Omega \frac{1}{\varepsilon}(1-\theta^*)(g_\beta(\mathbf{x}, \mathsf{u}^\varepsilon) - g_\beta(\mathbf{x}, \mathsf{u}^*))\,d\mathbf{x}.\end{aligned} \tag{2.16}$$

The first integral on the right hand side of (2.16) converges to

$$\int_\Omega \delta\theta(g_\alpha(\mathbf{x}, \mathsf{u}^*) - g_\beta(\mathbf{x}, \mathsf{u}^*) + l)\,d\mathbf{x},$$

due to $\lim_{\varepsilon\searrow 0}\frac{\theta^\varepsilon - \theta^*}{\varepsilon} = \delta\theta$ and Lemma 2.1, since u^ε strongly converges to u^* in $\mathrm{H}_0^1(\Omega; \mathbf{R}^d)$ (and, especially, in $\mathrm{L}^2(\Omega; \mathbf{R}^d)$). To obtain the limit of the second integral, we first apply the Lagrange mean value theorem. For a.e. $\mathbf{x} \in \Omega$ (more precisely, for any Lebesgue point $\mathbf{x}$ of all functions included in the following calculation),

$$\frac{g_\alpha(\mathbf{x}, \mathsf{u}^\varepsilon(\mathbf{x})) - g_\alpha(\mathbf{x}, \mathsf{u}^*(\mathbf{x}))}{\varepsilon} = \sum_{i=1}^m \frac{\partial g_\alpha(\mathbf{x}, \boldsymbol{\eta}^\varepsilon(\mathbf{x}))}{\partial u_i}\cdot\frac{u_i^\varepsilon(\mathbf{x}) - u_i^*(\mathbf{x})}{\varepsilon},$$

where $\boldsymbol{\eta}^\varepsilon(\mathbf{x})$ belongs to segment $[\mathsf{u}^\varepsilon(\mathbf{x}), \mathsf{u}^*(\mathbf{x})]$. Therefore, $\boldsymbol{\eta}^\varepsilon$ converges strongly to u^* in $\mathrm{L}^2(\Omega)$. Consequently,

$$\lim_{\varepsilon\searrow 0}\int_\Omega \frac{1}{\varepsilon}\theta^*(g_\alpha(\mathbf{x}, \mathsf{u}^\varepsilon(\mathbf{x})) - g_\alpha(\mathbf{x}, \mathsf{u}^*(\mathbf{x})))\,d\mathbf{x} = \int_\Omega \theta^*\sum_{i=1}^m \delta u_i\frac{\partial g_\alpha(\mathbf{x}, \mathsf{u}^*(\mathbf{x}))}{\partial u_i}\,d\mathbf{x}.$$

Similarly,

$$\lim_{\varepsilon\searrow 0}\int_\Omega \frac{1}{\varepsilon}(1-\theta^*)(g_\beta(\mathbf{x},\mathsf{u}^\varepsilon(\mathbf{x}))-g_\beta(\mathbf{x},\mathsf{u}^*(\mathbf{x})))\,d\mathbf{x}=\int_\Omega(1-\theta^*)\sum_{i=1}^m\delta u_i\frac{\partial g_\beta(\mathbf{x},\mathsf{u}^*(\mathbf{x}))}{\partial u_i}\,d\mathbf{x}.$$

By introducing adjoint states p_i^* as solutions of (2.12), for $i=1,\dots,m$, a weak formulation of the problems reads

$$\int_\Omega \mathbf{A}^*\nabla p_i^*(\mathbf{x})\cdot\nabla\psi(\mathbf{x})\,d\mathbf{x}=\int_\Omega\left(\theta^*\frac{\partial g_\alpha}{\partial u_i}(\mathbf{x},\mathsf{u}^*(\mathbf{x}))+(1-\theta^*)\frac{\partial g_\beta}{\partial u_i}(\mathbf{x},\mathsf{u}^*(\mathbf{x}))\right)\psi(\mathbf{x})\,d\mathbf{x},$$

for $\psi\in \mathrm{H}_0^1(\Omega)$. In particular, for $\psi=\delta u_i$, the necessary condition (2.14) transforms into

$$\int_\Omega (g_\alpha(\mathbf{x},\mathsf{u}^*(\mathbf{x}))-g_\beta(\mathbf{x},\mathsf{u}^*(\mathbf{x}))+l)\,\delta\theta(\mathbf{x})\,d\mathbf{x}-\int_\Omega\sum_{i=1}^m\delta\mathbf{A}(\mathbf{x})\nabla u_i^*(\mathbf{x})\cdot\nabla p_i^*(\mathbf{x})\,d\mathbf{x}\geq 0.$$

□

The main difficulty in further analysis of optimality condition (2.13) is that variations in θ and $\mathbf{A}$ are not independent. As the first step let us consider variations only in $\mathbf{A}$, taking $\delta\theta$ to be 0. Due to the convexity of $\mathcal{K}(\theta^*)$ (Remark 1.5), it is natural to take a segment in $\mathcal{K}(\theta^*)$ as an admissible path. Using this path, necessary conditions of optimality were derived in Allaire (2002), while Vrdoljak (2010) developed a variant of the optimality criteria method based on those conditions. Here, we choose another path: let us recall Lemma 1.12 on equivalence between inverse matrices of conductivity matrices from $\mathcal{K}(\theta^*)$ and $\tilde{\mathcal{K}}(\theta^*)$, where $\tilde{\mathcal{K}}(\theta^*)$ is the set of all matrices with eigenvalues $\nu_1,\dots,\nu_d$ satisfying

$$\nu_{\theta^*}^+\leq\ \nu_j\ \leq\nu_{\theta^*}^-\,,\quad j=1,\dots,d\,,\tag{2.17}$$
$$\sum_{j=1}^d\frac{1}{\alpha^{-1}-\nu_j}\ \leq\ \frac{1}{\alpha^{-1}-\nu_{\theta^*}^-}+\frac{d-1}{\alpha^{-1}-\nu_{\theta^*}^+}\,,\tag{2.18}$$
$$\sum_{j=1}^d\frac{1}{\nu_j-\beta^{-1}}\ \leq\ \frac{1}{\nu_{\theta^*}^--\beta^{-1}}+\frac{d-1}{\nu_{\theta^*}^+-\beta^{-1}}\,,\tag{2.19}$$

for $\nu_{\theta^*}^+=\dfrac{1}{\theta^*\alpha+(1-\theta^*)\beta}$ and $\nu_{\theta^*}^-=\dfrac{\theta^*}{\alpha}+\dfrac{1-\theta^*}{\beta}$. The set $\tilde{\mathcal{K}}(\theta^*)$ is also convex by Remark 1.5, and therefore we can take the admissible smooth path

$$\mathbf{A}^\varepsilon=(\varepsilon\mathbf{A}^{-1}+(1-\varepsilon)\mathbf{A}^{*-1})^{-1},\quad \mathbf{A}\in\mathcal{K}(\theta^*),$$

which represents a segment in $\tilde{\mathcal{K}}(\theta^*)$. Using the fact that

$$\frac{d}{d\varepsilon}\left(\mathbf{M}^{-1}(\varepsilon)\right)=-\mathbf{M}^{-1}(\varepsilon)\frac{d}{d\varepsilon}\left(\mathbf{M}(\varepsilon)\right)\mathbf{M}^{-1}(\varepsilon),$$

the admissible variation reads

$$\delta\mathbf{A} = \frac{d}{d\varepsilon}\left(\left(\varepsilon\mathbf{A}^{-1} + (1-\varepsilon)\mathbf{A}^{*-1}\right)^{-1}\right)\Big|_{\varepsilon=0^+} = \mathbf{A}^*(\mathbf{A}^{*-1} - \mathbf{A}^{-1})\mathbf{A}^*.$$

Substituting $\delta\mathbf{A}$ in the necessary condition of optimality (2.13), it follows (recall that $\mathbf{A}^*$ is symmetric)

$$\int_\Omega \sum_{i=1}^m (\mathbf{A}^{-1} - \mathbf{A}^{*-1})\mathbf{A}^*\nabla u_i^* \cdot \mathbf{A}^*\nabla p_i^* \, d\mathbf{x} \geq 0, \quad \mathbf{A} \in \mathcal{K}(\theta^*). \tag{2.20}$$

Let us denote $\boldsymbol{\sigma}_i^* = \mathbf{A}^*\nabla u_i^*$ and $\boldsymbol{\tau}_i^* = \mathbf{A}^*\nabla p_i^*$. Since $\mathbf{A} \in \mathcal{K}(\theta^*)$ is arbitrary, it follows

$$\sum_{i=1}^m \mathbf{A}^{-1}\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^* \geq \sum_{i=1}^m \mathbf{A}^{*-1}\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^*,$$

almost everywhere on Ω. On the contrary, if there exists $\tilde{\mathbf{A}} \in \mathcal{K}(\theta^*)$ and $\omega \subseteq \Omega$ of a nonzero measure such that

$$\sum_{i=1}^m \tilde{\mathbf{A}}^{-1}\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^* < \sum_{i=1}^m \mathbf{A}^{*-1}\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^*, \qquad \text{a.e. on } \omega,$$

then the monotonicity property of the integral implies

$$\int_\omega \sum_{i=1}^m (\tilde{\mathbf{A}}^{-1} - \mathbf{A}^{*-1})\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^* \, d\mathbf{x} < 0.$$

By defining a matrix $\mathbf{A}_1$ in a way that $\mathbf{A}_1 = \tilde{\mathbf{A}}$ on ω and $\mathbf{A}_1 = \mathbf{A}^*$ on $\Omega \setminus \omega$, which also belongs to the set $\mathcal{K}(\theta^*)$ we have

$$\int_\Omega \sum_{i=1}^m (\mathbf{A}_1^{-1} - \mathbf{A}^{*-1})\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^* \, d\mathbf{x} = \int_\omega \sum_{i=1}^m (\mathbf{A}_1^{-1} - \mathbf{A}^{*-1})\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^* \, d\mathbf{x} < 0,$$

which is a contradiction with (2.20).

Therefore, $\mathbf{A}^*$ is a solution of the minimization problem (a.e. on Ω)

$$\begin{cases} \sum_{i=1}^m \mathbf{A}^{-1}\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^* \to \min \\ \mathbf{A} \in \mathcal{K}(\theta^*), \end{cases} \tag{2.21}$$

which is a constrained minimization of a linear function. The above sum can be rewritten as

$$\sum_{i=1}^m \mathbf{A}^{-1}\boldsymbol{\sigma}_i^* \cdot \boldsymbol{\tau}_i^* = \mathbf{A}^{-1} : \sum_{i=1}^m \boldsymbol{\sigma}_i^* \otimes \boldsymbol{\tau}_i^* = \mathbf{A}^{-1} : \operatorname{Sym} \sum_{i=1}^m \boldsymbol{\sigma}_i^* \otimes \boldsymbol{\tau}_i^*,$$

where the last equality arises from symmetry of matrix $\mathbf{A}^{-1}$. By introducing a matrix function $\mathbf{N}^* = \mathrm{Sym} \sum_{i=1}^m \boldsymbol{\sigma}_i^* \otimes \boldsymbol{\tau}_i^*$, it follows

$$(\mathbf{A}^*)^{-1} : \mathbf{N}^* = \min_{\mathbf{A} \in \mathcal{K}(\theta^*)} \mathbf{A}^{-1} : \mathbf{N}^*, \quad \text{a.e. on } \Omega. \tag{2.22}$$

If $\mathbf{N}^* = 0$, any $\mathbf{A} \in \mathcal{K}(\theta^*)$ is optimal, otherwise we use the classical von Neumann result (von Neumann (1937)):

Theorem 2.5 For symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ with eigenvalues $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_d$ and $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_d$ respectively, it holds

$$\mathbf{A} : \mathbf{B} \leq \sum_{i=1}^d \alpha_i \beta_i,$$

with equality if and only if matrices $\mathbf{A}$ and $\mathbf{B}$ are simultaneously diagonalizable.

□

Let us denote the eigenvalues of symmetric matrices $\mathbf{N}^*$ and $\mathbf{A}^{-1}$ by $\eta_1^* \geq \eta_2^* \geq \ldots \geq \eta_d^*$ and $\nu_1 \leq \nu_2 \leq \cdots \leq \nu_d$, respectively. From Theorem 2.5 it follows

$$-\mathbf{A}^{-1} : \mathbf{N}^* \leq -\sum_{i=1}^d \nu_i \eta_i^*,$$

which implies

$$\mathbf{A}^{-1} : \mathbf{N}^* \geq \sum_{i=1}^d \nu_i \eta_i^* \geq \min_{\nu_i \in \mathcal{V}(\alpha,\beta;\theta^*)} \sum_{i=1}^d \nu_i \eta_i^*.$$

Therefore, from the condition of optimality (2.22) and Theorem 2.5, $(\mathbf{A}^*)^{-1}$ is simultaneously diagonalizable with $\mathbf{N}^*$, and its eigenvalues are optimal in the following problem:

$$\begin{cases} \displaystyle\sum_{j=1}^d \nu_j \eta_j^* \longrightarrow \min \\ (\nu_1, \nu_2, \ldots, \nu_d) \in \mathcal{V}(\alpha, \beta; \theta^*). \end{cases} \tag{2.23}$$

Remark 2.2. Clearly, the optimization problem (2.23), and thus (2.22), has a solution. This solution is not unique (Antonić & Vrdoljak (2011)) if and only if $\eta_{d-1}^* = \eta_d^* = 0$. In that case, the set of solution is infinite, one of them being $(\nu_{\theta^*}^+, \nu_{\theta^*}^+, \ldots, \nu_{\theta^*}^+, \nu_{\theta^*}^-)$, which corresponds to a simple laminate with the lamination direction from $\ker \mathbf{N}^*$ (see also the remark below).

Remark 2.3. If $\eta_d^* = 0$, the optimization problem (2.23) has a solution, which is a simple laminate. Indeed, for $\mathbf{A} \in \mathcal{K}(\theta^*)$ we have

$$\mathbf{A}^{-1} : \mathbf{N}^* = \sum_{i=1}^d \nu_i \eta_i^* = \sum_{i=1}^{d-1} \nu_i \eta_i^* \geq \sum_{i=1}^{d-1} \nu_{\theta^*}^+ \eta_i^*,$$

with equality achieved when $\mathbf{A}$ is a simple laminate with the lamination direction from

$\ker \mathbf{N}^*$ (an eigenvector of $\mathbf{N}^*$ corresponding to the eigenvalue $\eta_d^* = 0$). If $\eta_{d-1}^* \neq 0$ this solution is unique (Antonić & Vrdoljak (2011)), as already commented in the previous remark.

Now we take into account variations in θ and consider an admissible smooth path $\varepsilon \mapsto (\theta^\varepsilon, \mathbf{A}^\varepsilon)$ such that almost everywhere on Ω

$$(\mathbf{A}^\varepsilon)^{-1} : \mathbf{N}^* = g(\theta^\varepsilon, \mathbf{N}^*), \tag{2.24}$$

where function $g : [0,1] \times \mathrm{Sym}_d \to \mathbf{R}$ is defined by

$$g(\theta, \mathbf{N}) = \min_{\mathbf{A} \in \mathcal{K}(\theta)} (\mathbf{A}^{-1} : \mathbf{N}).$$

Since $\mathcal{K}(\theta^\varepsilon)$ is convex and $g(\theta^\varepsilon, \mathbf{N}^*) = \min\limits_{\mathbf{A} \in \mathcal{K}(\theta^\varepsilon)} (\mathbf{A}^{-1} : \mathbf{N}^*)$ a.e. on Ω, $\mathbf{A}^\varepsilon(\mathbf{x})$ belongs to the boundary of $\mathcal{K}(\theta^\varepsilon(\mathbf{x}))$, a.e. $\mathbf{x} \in \Omega$. Thus, $\mathbf{A}^\varepsilon(\mathbf{x})$ can be taken as a sequential laminate in the same directions with just a smooth variation of the proportions.

Since $\theta \mapsto g(\theta, \mathbf{N})$ is differentiable, as we shall see later, using variations $(\delta\theta, \delta\mathbf{A})$ generated by this smooth path, from (2.24) it follows

$$\delta\mathbf{A}^{-1} : \mathbf{N}^* = \frac{\partial g}{\partial \theta}(\theta^*, \mathbf{N}^*)\delta\theta \quad \text{a.e. on } \Omega,$$

and from the necessary condition of optimality (2.13) we get

$$\int_\Omega \delta\theta(\mathbf{x}) \left(g_\alpha(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) - g_\beta(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) + l + \frac{\partial g}{\partial \theta}(\theta^*(\mathbf{x}), \mathbf{N}^*(\mathbf{x})) \right) d\mathbf{x} \geq 0.$$

$\delta\theta$ can be chosen to vanish everywhere except on an arbitrary measurable subset of Ω, thus the inequality

$$\delta\theta(\mathbf{x}) \left(g_\alpha(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) - g_\beta(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) + l + \frac{\partial g}{\partial \theta}(\theta^*(\mathbf{x}), \mathbf{N}^*(\mathbf{x})) \right) \geq 0, \quad \text{a.e.}\, \mathbf{x} \in \Omega$$

is valid. Since $\mathrm{L}^\infty(\Omega; [0,1])$ is convex, an admissible variation can be taken as $\delta\theta = \theta - \theta^*$, for some $\theta \in \mathrm{L}^\infty(\Omega; [0,1])$. Therefore, if $\theta^*(\mathbf{x}) = 0$, then $\delta\theta(\mathbf{x}) \geq 0$, if $\theta^*(\mathbf{x}) = 1$, then $\delta\theta(\mathbf{x}) \leq 0$, while if $\theta^*(\mathbf{x}) \in \langle 0, 1 \rangle$, then $\delta\theta(\mathbf{x})$ can be positive and negative. This proves the following result.

Theorem 2.6 Let $(\theta^*, \mathbf{A}^*)$ be a local minimizer for the relaxation problem (2.9) with corresponding states u_i^* and adjoint states p_i^*. We define the symmetric matrix

$$\mathbf{N}^* := \mathrm{Sym} \sum_{i=1}^m \boldsymbol{\sigma}_i^* \otimes \boldsymbol{\tau}_i^*,$$

for $\boldsymbol{\sigma}_i^* := \mathbf{A}^*\nabla u_i^*$, $\boldsymbol{\tau}_i^* := \mathbf{A}^*\nabla p_i^*$, and function

$$R^*(\mathbf{x}) := g_\alpha(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) - g_\beta(\mathbf{x}, \mathsf{u}^*(\mathbf{x})) + l + \frac{\partial g}{\partial \theta}(\theta^*(\mathbf{x}), \mathbf{N}^*(\mathbf{x}))\,, \quad \text{a.e. } \mathbf{x} \in \Omega\,.$$

Then the optimal $\mathbf{A}^*$ satisfies

$$(\mathbf{A}^*)^{-1} : \mathbf{N}^* = \min_{\mathbf{A}\in\mathcal{K}(\theta^*)} \mathbf{A}^{-1} : \mathbf{N}^*, \quad \text{a.e. on } \Omega, \tag{2.25}$$

while the optimal θ^* satisfies (almost everywhere on Ω)

$$\begin{aligned} R^*(\mathbf{x}) > 0 &\implies \theta^*(\mathbf{x}) = 0\,, \\ R^*(\mathbf{x}) < 0 &\implies \theta^*(\mathbf{x}) = 1. \end{aligned} \tag{2.26}$$

□

Although we got existence results and necessary conditions of optimality for the relaxed formulation, exact solutions are obtained only for some simple domains and functionals (Burazin (2018), Burazin & Vrdoljak (2018), Casado-Díaz (2015*b*), Vrdoljak (2016)). For more complicated domains (or functionals) it is quite unlikely to find an analytic solution (Goodman et al. (1986)), which imposes a need for various numerical methods. One of them is the optimality criteria method, an iterative method based on the necessary conditions of optimality of the relaxed problem.

2.2 Optimality criteria method

A numerical method which provides a solution that satisfies the necessary condition of optimality for the optimal design problem is called the *optimality criteria method*. There are a numerous results on this method in structural engineering (see Bendsøe (1995), Rozvany (1989), and references therein). It is very effective when the optimality conditions are written in a simple manner and more efficient regarding computational time and required memory than some other methods (for example, gradient-based methods). Recently, Vrdoljak (2010) developed a variant of the optimality criteria method for the multiple state optimal design problems in conductivity setting, based on conditions of optimality derived in Allaire (2002). His method gives good convergence results in maximizing the conic sum of energy, while it fails for the minimization of the same functional. Therefore, in this dissertation we present another variant of the optimality criteria method suitable for those problems.

2.2.1 The algorithm

Another advantage of the optimality criteria method is that the algorithm is quite simple. The principle of the method is to iteratively compute the state function and design parameters, using conditions of optimality. Since we introduced new optimality conditions given by Theorem 2.6, we are able to write down a new variant of the optimality criteria method (Burazin et al. (2018)) listed below.

Algorithm 2.7 Take some initial $(\theta^0, \mathbf{A}^0) \in \mathcal{A}$. For $k \geq 0$:

(1) Calculate the solution u_i^k, $i = 1, \dots, m$, of

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^k \nabla u_i^k) = f_i \quad \text{in } \Omega \\ u_i^k \in \mathrm{H}_0^1(\Omega). \end{cases}$$

(2) Calculate the solution p_i^k, $i = 1, \dots, m$, of

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^k \nabla p_i^k) = \theta^k \dfrac{\partial g_\alpha}{\partial u_i}(\cdot, \mathsf{u}^k) + (1-\theta^k)\dfrac{\partial g_\beta}{\partial u_i}(\cdot, \mathsf{u}^k) \quad \text{in } \Omega \\ p_i^k \in \mathrm{H}_0^1(\Omega),\ \mathsf{u}^k = (u_1^k, \dots, u_m^k) \end{cases}$$

and define $\boldsymbol{\sigma}_i^k := \mathbf{A}^k \nabla u_i^k$, $\boldsymbol{\tau}_i^k := \mathbf{A}^k \nabla p_i^k$ and $\mathbf{N}^k := \mathrm{Sym} \sum_{i=1}^m (\boldsymbol{\sigma}_i^k \otimes \boldsymbol{\tau}_i^k)$.

If $\mathbf{N}^k(\mathbf{x}) = \mathbf{0}$, for $\mathbf{x} \in \Omega$, leave the old data for the next iteration of $\theta^{k+1}(\mathbf{x})$ and $\mathbf{A}^{k+1}(\mathbf{x})$. Else, do steps 3 and 4.

(3) For $\mathbf{x} \in \Omega$, let $\theta^{k+1}(\mathbf{x}) \in [0,1]$ be a zero of the function

$$\theta \mapsto R^k(\theta, \mathbf{x}) := g_\alpha(\mathbf{x}, \mathsf{u}^k(\mathbf{x})) - g_\beta(\mathbf{x}, \mathsf{u}^k(\mathbf{x})) + l + \frac{\partial g}{\partial \theta}(\theta, \mathbf{N}^k(\mathbf{x}))\,, \tag{2.27}$$

and if a zero doesn't exist, take 0 (or 1) if the function is positive (or negative) on $[0,1]$.

(4) For $\mathbf{x} \in \Omega$, let $\mathbf{A}^{k+1}(\mathbf{x})$ be a minimizer in $g(\theta^{k+1}(\mathbf{x}), \mathbf{N}^k(\mathbf{x})) = \min\limits_{\mathbf{A} \in \mathcal{K}(\theta^{k+1})} (\mathbf{A}^{-1} : \mathbf{N}^k)$. Since this minimizer is not unique in general, let us be more precise with this step: first diagonalize the matrix $\mathbf{N}^k(\mathbf{x})$, then find the solution $\nu_1^{k+1} \leq \dots \leq \nu_d^{k+1}$ of the minimization problem ($\eta_1^k \geq \eta_2^k \geq \ldots \geq \eta_d^k$ are eigenvalues of $\mathbf{N}^k$)

$$\begin{cases} \displaystyle\sum_{j=1}^d \nu_j \eta_j^k \longrightarrow \min \\ \nu_{\theta^{k+1}}^+ \leq \nu_j\,, \\ \displaystyle\sum_{j=1}^d \frac{1}{\alpha^{-1} - \nu_j} \leq \frac{1}{\alpha^{-1} - \nu_{\theta^{k+1}}^-} + \frac{d-1}{\alpha^{-1} - \nu_{\theta^{k+1}}^+}\,, \\ \displaystyle\sum_{j=1}^d \frac{1}{\nu_j - \beta^{-1}} \leq \frac{1}{\nu_{\theta^{k+1}}^- - \beta^{-1}} + \frac{d-1}{\nu_{\theta^{k+1}}^+ - \beta^{-1}}\,, \end{cases} \tag{2.28}$$

and determine eigenvalues $\lambda_i^{k+1} = (\nu_i^{k+1})^{-1}$ of the matrix $\mathbf{A}^{k+1}$. When the solution is not unique (i.e. when $\eta_{d-1}^k = \eta_d^k = 0$; see Remark 2.2), take $\lambda_i^{k+1} = \lambda_{\theta^{k+1}}^+$, $i = 1, \ldots, d-1$ and $\lambda_d^{k+1} = \lambda_{\theta^{k+1}}^-$.

Calculate the next $\mathbf{A}^{k+1} = \mathbf{Q}\,\mathrm{diag}(\lambda_1^{k+1}, \ldots, \lambda_d^{k+1})\mathbf{Q}^\top$, where $\mathbf{Q}$ is a matrix whose columns are eigenvectors of the matrix $\mathbf{N}^k$.

Remark 2.4. In the case when $\eta_d^k = 0$ in the fourth step of the above algorithm, the next iteration of $\mathbf{A}^{k+1}$ is a simple laminate with the lamination direction from $\ker \mathbf{N}^k$ (see Remark 2.3) However, when additionally $\eta_{d-1}^k = 0$, notice that this laminate is not uniquely determined, as $\dim \ker \mathbf{N}^k \geq 2$, and thus the lamination direction is not unique. In addition, there exist solutions other than simple laminates because it is only needed to have $\lambda_1^{k+1} = \cdots = \lambda_{d-2}^{k+1} = \lambda_{\theta^k}^+$, while λ_{d-1}^{k+1} and λ_d^{k+1} can be arbitrary.

Remark 2.5. An appropriate criterion for stopping the algorithm could be when $|J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ is small enough.

Remark 2.6. In Allaire (2002), this variant of the optimality criteria method was developed for solving single state self-adjoint minimization problems, i.e. when $u = \pm p$. In this case, taking a pair $(\theta^0, \mathbf{A}^0) \in \mathcal{A}$ as initial design, for $k \geq 0$ the algorithm from Allaire (2002) iteratively computes solution $u^k = p^k$ of the state equation (1.1) with the design parameters $(\theta^k, \mathbf{A}^k)$, and then updates these parameters by taking $\mathbf{A}^{k+1}$ as a simple laminate with the lamination direction orthogonal to $\boldsymbol{\sigma}^k = \mathbf{A}^k \nabla u^k$, and θ^{k+1} as the root of the equation

$$l + (\beta - \alpha)\frac{|\boldsymbol{\sigma}^k|^2}{(\lambda_\theta^+)^2} = 0. \tag{2.29}$$

Algorithm 2.7 is actually a generalisation of the above mentioned algorithm. Indeed, for $m = 1$, let $(\theta^k, \mathbf{A}^k)$ be a design obtained in k-th step of Algorithm 2.7, for the same initial pair $(\theta^0, \mathbf{A}^0)$. Matrix $\mathbf{N}^k = \boldsymbol{\sigma}^k \otimes \boldsymbol{\sigma}^k$ has one eigenvalue equal to $|\boldsymbol{\sigma}^k|^2$, which corresponds to the eigenvector $\boldsymbol{\sigma}^k$, while all other eigenvalues are equal to zero. This implies (see remarks 2.3 and 2.4) that the algorithm gives $\mathbf{A}^{k+1}$ as a simple laminate with a lamination direction from $\ker \mathbf{N}^k$, and thus orthogonal to $\boldsymbol{\sigma}^k$. The function $g(\theta, \mathbf{N}^k)$ in this case is given by $g(\theta, \mathbf{N}^k) = \frac{|\boldsymbol{\sigma}^k|^2}{\lambda_\theta^+}$, while $R^k(\theta, \mathbf{x}) = l + (\beta - \alpha)\frac{|\boldsymbol{\sigma}^k|^2}{(\lambda_\theta^+)^2}$. Therefore, θ^{k+1} as a zero of the function R^k, coincides with the root of equation (2.29), which implies that Algorithm 2.7 for $m = 1$ coincides with the algorithm developed in Allaire (2002).

In order to implement Algorithm 2.7, we need an explicit calculation of a zero point, if it exists, of the function $\theta \mapsto R^k(\theta, \mathbf{x})$. In the sequel, we shall present explicit formulae for the partial derivative $\frac{\partial g}{\partial \theta}$ for the general (multi-state) case.

Let us first consider the two-dimensional case. As commented, the minimization over $\mathcal{K}(\theta)$ in the definition of function g can be expressed equivalently by minimization over eigenvalues:

$$g(\theta, \mathbf{N}) = \min_{\boldsymbol{\nu} \in \mathcal{V}(\alpha,\beta;\theta)} \sum_{j=1}^{d} \nu_j \eta_j \,,$$

where η_j are the eigenvalues of the symmetric matrix $\mathbf{N}$.

In the two-dimensional case the set $\mathcal{V}(\alpha, \beta; \theta)$ is equal to the set $\Lambda\left(\frac{1}{\beta}, \frac{1}{\alpha}; 1-\theta\right)$ (see Remark 1.4). This remark can be used to calculate g and its partial derivative over θ on the basis of (Allaire 2002, Lemma 3.2.17), as presented in the next theorem.

Theorem 2.8 In the case $d = 2$, for given $\theta \in [0, 1]$ and a symmetric matrix $\mathbf{N}$ with eigenvalues $\eta_1 \geq \eta_2$, we have

A. If $\eta_2 > 0$, then for $\theta^A := \left(\alpha\frac{\sqrt{\eta_1}}{\sqrt{\eta_2}} - \beta\right)\frac{1}{\alpha-\beta}$ it holds

$$\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \begin{cases} \frac{1}{\beta}\left(\beta^2-\alpha^2\right)\left(\frac{\sqrt{\eta_1}+\sqrt{\eta_2}}{\theta(\alpha-\beta)+\beta+\alpha}\right)^2, & \theta < \theta^A \\ \frac{(\beta-\alpha)\,\eta_1}{(\theta(\alpha-\beta)+\beta)^2} + \eta_2\left(\frac{1}{\alpha}-\frac{1}{\beta}\right), & \theta \geq \theta^A \end{cases}.$$

B. If $\eta_1 < 0$, then for $\theta^B := \left(\frac{\sqrt{-\eta_1}}{\sqrt{-\eta_2}} - 1\right)\frac{\beta}{\alpha-\beta}$ it holds

$$\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \begin{cases} -\frac{1}{\alpha}\left(\beta^2-\alpha^2\right)\left(\frac{\sqrt{-\eta_1}+\sqrt{-\eta_2}}{\theta(\alpha-\beta)+2\beta}\right)^2, & \theta > \theta^B \\ \frac{(\beta-\alpha)\,\eta_1}{(\theta(\alpha-\beta)+\beta)^2} + \eta_2\left(\frac{1}{\alpha}-\frac{1}{\beta}\right), & \theta \leq \theta^B \end{cases}.$$

C. If $\eta_1 \geq 0$ and $\eta_2 \leq 0$, then

$$\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \frac{(\beta-\alpha)\,\eta_1}{(\theta(\alpha-\beta)+\beta)^2} + \eta_2\left(\frac{1}{\alpha}-\frac{1}{\beta}\right).$$

Proof. In the proof we emphasize parameters α and β in the set $\mathcal{K}(\theta)$ by denoting it by $\mathcal{K}(\alpha, \beta;\, \theta)$, for given $\theta \in [0,\, 1]$. For dimension $d = 2$, condition $\mathbf{A} \in \mathcal{K}(\alpha, \beta;\, \theta)$ can be equivalently expressed as $\mathbf{A}^{-1} \in \mathcal{K}\left(\frac{1}{\beta}, \frac{1}{\alpha};\, 1-\theta\right)$. Now it follows

$$g(\theta,\, \mathbf{N}) = \min_{\mathbf{A}\in\mathcal{K}(\alpha,\beta;\,\theta)} \mathbf{A}^{-1} : \mathbf{N} = -\max_{\mathbf{A}^{-1}\in\mathcal{K}\left(\frac{1}{\beta},\frac{1}{\alpha};\,1-\theta\right)} \mathbf{A}^{-1} : (-\mathbf{N}) = -f_{1/\beta}^{1/\alpha}(1-\theta,\, -\mathbf{N}),$$

where, for $0 < \gamma < \delta$, function $f_\gamma^\delta : [0,\, 1] \times \mathrm{Sym}_d \longrightarrow \mathbf{R}$ is defined in (Allaire 2002, Theorem 3.2.14),

$$f_\gamma^\delta(\theta, \mathbf{M}) = \max_{\mathbf{A}\in\mathcal{K}(\gamma,\,\delta;\,\theta)} \mathbf{A} : \mathbf{M}\,.$$

Furthermore,

$$\frac{\partial g}{\partial \theta}(\theta,\, \mathbf{N}) = \frac{\partial f_{1/\beta}^{1/\alpha}}{\partial \theta}(1-\theta,\, -\mathbf{N}).$$

Therefore, one can use the formula for f_α^β given in (Allaire 2002, Lemma 3.2.17) to obtain $\frac{\partial g}{\partial \theta}$. ☐

In this case ($d = 2$), the function $\theta \mapsto R^k(\theta, \mathbf{x})$ introduced in (2.27) is strictly monotone on $[0, 1]$, for almost every $\mathbf{x} \in \Omega$, so its zero point (if it exists) is unique. Moreover, by formulae presented in Theorem 2.8, the zero point can be calculated explicitly, as a zero of a quadratic equation. For example, if the eigenvalues of matrix $\mathbf{N}^k$ fits the case A above, then the function $\theta \mapsto R^k(\theta, \mathbf{x})$ is strictly increasing. Therefore, one should simply check signs of $R^k(\theta, \mathbf{x})$ for $\theta \in \{0, 1\}$ (and $\theta = \theta_A$, if $0 < \theta^A < 1$) to locate the zero point (if it exists), and solve the corresponding quadratic equation for θ.

In the three-dimensional case, the situation is more tedious, and we shall begin by solving the minimization problem (2.23).

Theorem 2.9 In the case $d = 3$, let $0 < \theta < 1$ and $\eta_1 \geq \eta_2 \geq \eta_3$ be given. Then the minimization problem

$$\begin{cases} \nu_1\eta_1 + \nu_2\eta_2 + \nu_3\eta_3 \longrightarrow \min \\ \\ (\nu_1, \nu_2, \nu_3) \in \mathcal{V}(\alpha, \beta; \theta), \end{cases} \tag{2.30}$$

has a solution ν^* as follows:

I. If $\left(\eta_3 < 0 \text{ and } \eta_2 \geq \eta_3 \left(\frac{1-\alpha\nu_\theta^-}{1-\alpha\nu_\theta^+}\right)^2\right)$ or $\left(\eta_3 \geq 0 \text{ and } \eta_2 \geq \eta_3 \left(\frac{\beta\nu_\theta^- - 1}{\beta\nu_\theta^+ - 1}\right)^2\right)$, then $\nu^* = \left(\nu_\theta^+,\, \nu_\theta^+,\, \nu_\theta^-\right)$ is optimal.

II. Let $\eta_2 < \eta_3 \left(\frac{1-\alpha\nu_\theta^-}{1-\alpha\nu_\theta^+}\right)^2$ (this is possible only if $\eta_2 < 0$).

1. If $\eta_1 \geq 0$ or else if $\sqrt{-\eta_2} + \sqrt{-\eta_3} \geq \sqrt{-\eta_1}\left(1 + \frac{1-\alpha\nu_\theta^+}{1-\alpha\nu_\theta^-}\right)$ then $\nu^* = \left(\nu_\theta^+,\, \nu_2,\, \nu_3\right)$ is optimal, where

$$\nu_i = \frac{1}{\alpha} - \frac{1}{\sqrt{-\eta_i}} \frac{\sqrt{-\eta_2} + \sqrt{-\eta_3}}{c_1(\theta)}, \qquad i = 2, 3; \tag{2.31}$$

with $c_1(\theta) = \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{1}{\alpha^{-1} - \nu_\theta^+}$.

2. Otherwise, if $\eta_1 < 0$ and $\sqrt{-\eta_2} + \sqrt{-\eta_3} < \sqrt{-\eta_1}\left(1 + \frac{1-\alpha\nu_\theta^+}{1-\alpha\nu_\theta^-}\right)$ then $\nu^* = (\nu_1,\, \nu_2,\, \nu_3)$ is optimal, where

$$\nu_i = \frac{1}{\alpha} - \frac{1}{\sqrt{-\eta_i}} \frac{\sqrt{-\eta_1} + \sqrt{-\eta_2} + \sqrt{-\eta_3}}{c_2(\theta)}, \qquad i = 1, 2, 3; \tag{2.32}$$

with $c_2(\theta) = \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{2}{\alpha^{-1} - \nu_\theta^+}$.

III. Let $\eta_2 < \eta_3 \left(\frac{\beta\nu_\theta^- - 1}{\beta\nu_\theta^+ - 1}\right)^2$ (this is possible only if $\eta_3 > 0$).

1. If $\sqrt{\eta_2}+\sqrt{\eta_3} \le \sqrt{\eta_1}\left(1+\frac{\beta\nu_\theta^+-1}{\beta\nu_\theta^--1}\right)$ then $\nu^*=\left(\nu_\theta^+,\,\nu_2,\,\nu_3\right)$ is optimal, where

$$\nu_i = \frac{1}{\beta}+\frac{1}{\sqrt{\eta_i}}\frac{\sqrt{\eta_2}+\sqrt{\eta_3}}{d_1(\theta)}, \qquad i=2,3; \tag{2.33}$$

with $d_1(\theta)=\frac{1}{\nu_\theta^- -\beta^{-1}}+\frac{1}{\nu_\theta^+-\beta^{-1}}$.

2. If $\sqrt{\eta_2}+\sqrt{\eta_3} > \sqrt{\eta_1}\left(1+\frac{\beta\nu_\theta^+-1}{\beta\nu_\theta^--1}\right)$ then $\nu^*=(\nu_1,\,\nu_2,\,\nu_3)$ is optimal, where

$$\nu_i = \frac{1}{\beta}+\frac{1}{\sqrt{\eta_i}}\frac{\sqrt{\eta_1}+\sqrt{\eta_2}+\sqrt{\eta_3}}{d_2(\theta)}, \qquad i=1,2,3; \tag{2.34}$$

with $d_2(\theta)=\frac{1}{\nu_\theta^- -\beta^{-1}}+\frac{2}{\nu_\theta^+-\beta^{-1}}$.

Proof. Note that due to the symmetry of the set $\mathcal{V}(\alpha,\beta;\theta)$ in $\nu_1,\,\nu_2,\,\nu_3$, we can conclude that a minimum point satisfies $\nu_\theta^+\le\nu_1\le\nu_2\le\nu_3$. Moreover, by observing that we are minimizing a linear function over a convex set, the optimal point belongs to the boundary of the set $\mathcal{V}(\alpha,\beta;\theta)$ and conversely, every boundary point of $\mathcal{V}(\alpha,\beta;\theta)$ can be obtained as a solution of (2.30) for some η_1, η_2 and η_3. In addition, if $\eta_1\ge 0$ and $\eta_2=\eta_3=0$, the problem (2.30) has a non-unique solution, one of them being a simple laminate $\nu^*=(\nu_\theta^+,\nu_\theta^+,\nu_\theta^-)$. Otherwise, there is a unique minimizer which we find by solving the Karush-Kuhn-Tucker (KKT) system (Karush (1939), Kuhn & Tucker (1951)). We already eliminated flat parts of the boundary of the set $\mathcal{V}(\alpha,\beta;\theta)$ (non-uniqueness of the solution appears here), so we have to analyze the rest of the boundary consisting precisely of: simple laminates, second and third order sequential laminates with matrix material α, and second and third order sequential laminates with matrix material β (see Figure 1.1). These five cases correspond exactly to cases I, II.1, II.2, III.1, and III.2 of Theorem 2.9.

We start with part I of the theorem. Assume that minimizer $\nu^*=(\nu_1,\,\nu_2,\,\nu_3)$ belongs to the part of the boundary of $\mathcal{V}(\alpha,\beta;\theta)$ corresponding to a simple laminate, i.e. $\nu^*=(\nu_\theta^+,\,\nu_\theta^+,\,\nu_\theta^-)$. We shall derive conditions on η_1 , η_2 and η_3 which ensure that optimal ν^* belongs to this part of the boundary. In this case, the KKT system reads:

$$\eta_1 = \frac{-a_1}{(\alpha^{-1}-\nu_\theta^+)^2}+\frac{a_2}{(\nu_\theta^+-\beta^{-1})^2}+a_3, \tag{2.35}$$

$$\eta_2 = \frac{-a_1}{(\alpha^{-1}-\nu_\theta^+)^2}+\frac{a_2}{(\nu_\theta^+-\beta^{-1})^2}+a_4, \tag{2.36}$$

$$\eta_3 = \frac{-a_1}{(\alpha^{-1}-\nu_\theta^-)^2}+\frac{a_2}{(\nu_\theta^--\beta^{-1})^2}, \tag{2.37}$$

for some nonnegative multipliers a_i, $i=1,\ldots,4$. Additionally, if $\eta_3<0$, we subtract equation (2.37) from (2.36), previously multiplied by $(\alpha^{-1}-\nu_\theta^-)^2$ and $(\alpha^{-1}-\nu_\theta^+)^2$, respectively, and get

$$\eta_2(\alpha^{-1}-\nu_\theta^+)^2-\eta_3(\alpha^{-1}-\nu_\theta^-)^2=a_2\left[\left(\frac{\alpha^{-1}-\nu_\theta^+}{\nu_\theta^+-\beta^{-1}}\right)^2-\left(\frac{\alpha^{-1}-\nu_\theta^-}{\nu_\theta^--\beta^{-1}}\right)^2\right]+a_4(\alpha^{-1}-\nu_\theta^+)^2.$$

A simple calculation verifies $\left(\frac{\alpha^{-1}-\nu_\theta^+}{\nu_\theta^+-\beta^{-1}}\right)^2-\left(\frac{\alpha^{-1}-\nu_\theta^-}{\nu_\theta^--\beta^{-1}}\right)^2\geq 0$, and the nonnegativity of a_2 and a_4 leads to the condition

$$\eta_2\geq\eta_3\left(\frac{1-\alpha\nu_\theta^-}{1-\alpha\nu_\theta^+}\right)^2,$$

which completes the first set of conditions for part I. On the other hand, if $\eta_3\geq 0$, we subtract equation (2.37) from (2.36), previously multiplied by $(\nu_\theta^- - \beta^{-1})^2$ and $(\nu_\theta^+ - \beta^{-1})^2$, respectively, and using an analogous procedure, we obtain

$$\eta_2\geq\eta_3\left(\frac{\beta\nu_\theta^- -1}{\beta\nu_\theta^+ -1}\right)^2,$$

which proves part I.

Let us now suppose that the minimizer $\nu^*=(\nu_1,\,\nu_2,\,\nu_3)$ belongs to the part of boundary of $\mathcal{V}(\alpha,\beta;\theta)$ corresponding to second order sequential laminates with matrix material α, described by

$$\nu_1 = \nu_\theta^+\,, \tag{2.38}$$

$$\nu_2,\,\nu_3 > \nu_\theta^+\,, \tag{2.39}$$

$$\sum_{j=1}^3\frac{1}{\alpha^{-1}-\nu_j} = \frac{1}{\alpha^{-1}-\nu_\theta^-}+\frac{2}{\alpha^{-1}-\nu_\theta^+}\,, \tag{2.40}$$

$$\sum_{j=1}^3\frac{1}{\nu_j-\beta^{-1}} < \frac{1}{\nu_\theta^- -\beta^{-1}}+\frac{2}{\nu_\theta^+ -\beta^{-1}}\,. \tag{2.41}$$

Again, we shall derive conditions on η_1 , η_2 and η_3 which ensure that an optimal ν^* belongs to this part of the boundary, and calculate the optimal ν^* in terms of η_1 , η_2 and η_3. The KKT system in this case is as follows:

$$\begin{aligned}\eta_1 &= \frac{-a_1}{(\alpha^{-1}-\nu_1)^2}+a_3,\\ \eta_2 &= \frac{-a_1}{(\alpha^{-1}-\nu_2)^2},\\ \eta_3 &= \frac{-a_1}{(\alpha^{-1}-\nu_3)^2},\end{aligned}$$

for some nonnegative multipliers a_1 and a_3. From the argument made at the beginning of the proof, we conclude $a_1 > 0$, implying that $\eta_2, \eta_3 < 0$ and

$$\frac{1}{\alpha^{-1} - \nu_i} = \sqrt{\frac{-\eta_i}{a_1}}, \qquad i = 2, 3, \tag{2.42}$$

which together with (2.40) gives

$$\sqrt{a_1} = \frac{\sqrt{-\eta_2} + \sqrt{-\eta_3}}{c_1(\theta)}, \quad \text{where } c_1(\theta) = \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{1}{\alpha^{-1} - \nu_\theta^+}. \tag{2.43}$$

Inserting this into (2.42), one obtains formula (2.31). It remains to identify under which conditions on η_1, η_2 and η_3, condition (2.39) is satisfied, with $\eta_1 = \frac{-a_1}{(\alpha^{-1} - \nu_1)^2} + a_3$, for $a_1,\ a_3 \geq 0$. A simple calculation gives that the condition $\nu_\theta^+ < \nu_2$ is equivalent to

$$\sqrt{-\eta_3} < \sqrt{-\eta_2}\, \frac{1 - \alpha\nu_\theta^+}{1 - \alpha\nu_\theta^-}, \tag{2.44}$$

while the condition $a_3 \geq 0$ leads to

$$\eta_1 + \frac{a_1}{(\alpha^{-1} - \nu_1)^2} \geq 0.$$

The above inequality is trivially satisfied if $\eta_1 \geq 0$, while if $\eta_1 < 0$, from (2.43), using (2.38), it is equivalent to the inequality

$$\sqrt{-\eta_2} + \sqrt{-\eta_3} \geq \sqrt{-\eta_1}\left(1 + \frac{1 - \alpha\nu_\theta^+}{1 - \alpha\nu_\theta^-}\right). \tag{2.45}$$

This proves part II.1 of the theorem. To prove part II.2, assume that the minimizer $\nu^* = (\nu_1,\ \nu_2,\ \nu_3)$ belongs to the part of boundary of $\mathcal{V}(\alpha, \beta; \theta)$ corresponding to third order sequential laminates with matrix material α, described by

$$\nu_1,\ \nu_2,\ \nu_3 \quad > \quad \nu_\theta^+\,, \tag{2.46}$$

$$\sum_{j=1}^{3} \frac{1}{\alpha^{-1} - \nu_j} \quad = \quad \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{2}{\alpha^{-1} - \nu_\theta^+}\,, \tag{2.47}$$

$$\sum_{j=1}^{3} \frac{1}{\nu_j - \beta^{-1}} \quad < \quad \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{2}{\nu_\theta^+ - \beta^{-1}}\,. \tag{2.48}$$

KKT system in this case reads

$$\begin{aligned} \eta_1 &= \frac{-a_1}{(\alpha^{-1} - \nu_1)^2}, \\ \eta_2 &= \frac{-a_1}{(\alpha^{-1} - \nu_2)^2}, \\ \eta_3 &= \frac{-a_1}{(\alpha^{-1} - \nu_3)^2}, \end{aligned}$$

where again $a_1 > 0$, which implies $\eta_1,\ \eta_2,\ \eta_3 < 0$ and

$$\frac{1}{\alpha^{-1} - \nu_i} = \sqrt{\frac{-\eta_i}{a_1}}, \qquad i = 1, 2, 3. \tag{2.49}$$

Inserting (2.49) into (2.47) we get

$$\sqrt{a_1} = \frac{\sqrt{-\eta_1} + \sqrt{-\eta_2} + \sqrt{-\eta_3}}{c_2(\theta)}, \quad \text{where } c_2(\theta) = \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{2}{\alpha^{-1} - \nu_\theta^+}. \tag{2.50}$$

Formula (2.32) is obtained by inserting (2.50) into (2.49). The condition $\nu_\theta^+ < \nu_1$ is equivalent to

$$\sqrt{-\eta_2} + \sqrt{-\eta_3} < \sqrt{-\eta_1}\left(1 + \frac{1 - \alpha\nu_\theta^+}{1 - \alpha\nu_\theta^-}\right), \tag{2.51}$$

which concludes part II. of the theorem.

Let us now suppose that the minimizer $\nu^* = (\nu_1,\ \nu_2,\ \nu_3)$ belongs to the part of boundary of $\mathcal{V}(\alpha, \beta; \theta)$ corresponding to second order sequential laminates with matrix material β, described by

$$\nu_1 = \nu_\theta^+\,, \tag{2.52}$$

$$\nu_2,\ \nu_3 > \nu_\theta^+\,, \tag{2.53}$$

$$\sum_{j=1}^{3} \frac{1}{\alpha^{-1} - \nu_j} < \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{2}{\alpha^{-1} - \nu_\theta^+}\,, \tag{2.54}$$

$$\sum_{j=1}^{3} \frac{1}{\nu_j - \beta^{-1}} = \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{2}{\nu_\theta^+ - \beta^{-1}}\,. \tag{2.55}$$

For this case, the KKT system reads:

$$\begin{aligned} \eta_1 &= \frac{a_2}{(\nu_1 - \beta^{-1})^2} + a_3, \\ \eta_2 &= \frac{a_2}{(\nu_1 - \beta^{-1})^2}, \\ \eta_3 &= \frac{a_2}{(\nu_1 - \beta^{-1})^2}, \end{aligned}$$

for some nonnegative multipliers a_2 and a_3. As before, we conclude $a_2 > 0$, implying that $\eta_2, \eta_3 > 0$ and

$$\frac{1}{\nu_i - \beta^{-1}} = \sqrt{\frac{\eta_i}{a_2}}, \qquad i = 2, 3, \tag{2.56}$$

which together with (2.55) gives

$$\sqrt{a_2} = \frac{\sqrt{\eta_2} + \sqrt{\eta_3}}{d_1(\theta)}, \quad \text{where } d_1(\theta) = \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{1}{\nu_\theta^+ - \beta^{-1}}. \tag{2.57}$$

Inserting the above into (2.56), one gets formula (2.33). It remains to find conditions on η_1, η_2 and η_3, with $\eta_1 = \frac{a_2}{\nu_1 - \beta^{-1})^2} + a_3$, for $a_2,\, a_3 \geq 0$, such that the condition (2.53) is satisfied. A simple calculation gives that the condition $\nu_\theta^+ < \nu_2$ is equivalent to

$$\sqrt{\eta_2} < \sqrt{\eta_3}\, \frac{\beta\nu_\theta^- - 1}{\beta\nu_\theta^+ - 1}. \tag{2.58}$$

From (2.57), using (2.52), the condition $a_3 \geq 0$ is equivalent to the inequality

$$\sqrt{\eta_2} + \sqrt{\eta_3} \leq \sqrt{\eta_1}\left(1 + \frac{\beta\nu_\theta^+ - 1}{\beta\nu_\theta^- - 1}\right). \tag{2.59}$$

At last, suppose that the minimizer $\nu^* = (\nu_1,\, \nu_2,\, \nu_3)$ belongs to the part of boundary of $\mathcal{V}(\alpha, \beta; \theta)$ corresponding to third order sequential laminates with matrix material β, described by

$$\begin{aligned} \nu_1,\, \nu_2,\, \nu_3 \quad &> \quad \nu_\theta^+\,, && (2.60)\\ \sum_{j=1}^{3} \frac{1}{\alpha^{-1} - \nu_j} \quad &< \quad \frac{1}{\alpha^{-1} - \nu_\theta^-} + \frac{2}{\alpha^{-1} - \nu_\theta^+}\,, && (2.61)\\ \sum_{j=1}^{3} \frac{1}{\nu_j - \beta^{-1}} \quad &= \quad \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{2}{\nu_\theta^+ - \beta^{-1}}\,. && (2.62) \end{aligned}$$

KKT system in this case is the following

$$\begin{aligned} \eta_1 &= \frac{a_2}{(\nu_1 - \beta^{-1})^2},\\ \eta_2 &= \frac{a_2}{(\nu_2 - \beta^{-1})^2},\\ \eta_3 &= \frac{a_2}{(\nu_3 - \beta^{-1})^2}, \end{aligned}$$

where $a_2 > 0$, which implies $\eta_1,\, \eta_2,\, \eta_3 > 0$ and

$$\frac{1}{\nu_i - \beta^{-1}} = \sqrt{\frac{\eta_i}{a_2}}, \qquad i = 1, 2, 3. \tag{2.63}$$

Inserting (2.63) into (2.62) we get

$$\sqrt{a_2} = \frac{\sqrt{\eta_1} + \sqrt{\eta_2} + \sqrt{\eta_3}}{d_2(\theta)}, \quad \text{where } d_2(\theta) = \frac{1}{\nu_\theta^- - \beta^{-1}} + \frac{2}{\nu_\theta^+ - \beta^{-1}}. \tag{2.64}$$

Formula (2.34) is obtained by inserting (2.64) into (2.63). The inequality $\nu_\theta^+ < \nu_1$ is equivalent to

$$\sqrt{\eta_2} + \sqrt{\eta_3} > \sqrt{\eta_1}\left(1 + \frac{\beta\nu_\theta^+ - 1}{\beta\nu_\theta^- - 1}\right), \tag{2.65}$$

and the proof is completed.

□

Before providing the function $g(\theta, \mathbf{N})$ and its derivatives, let us rewrite Theorem 2.9 in a more convenient way for implementation on a computer.

Corollary 2.10 In the case $d = 3$, for given $\eta_1 \geq \eta_2 \geq \eta_3$ and $0 < \theta < 1$, one can calculate the minimum point $\nu^* = (\nu_1,\, \nu_2,\, \nu_3)$ for (2.30) in the following way:

A If $\eta_3 = 0$, then the optimal point is $\nu^* = (\nu_\theta^+,\, \nu_\theta^+,\, \nu_\theta^-)$.

B Else if $\eta_3 > 0$, then calculate ν_1 by formula (2.34).

- If $\nu_\theta^+ < \nu_1$, then both ν_2 and ν_3 are given by (2.34).
- Else, calculate ν_2 by formula (2.33).
  - If $\nu_\theta^+ < \nu_2$, then $\nu_1 = \nu_\theta^+$ and ν_3 is given by (2.33).
  - Else $\nu^* = (\nu_\theta^+,\, \nu_\theta^+,\, \nu_\theta^-)$.

C Else ($\eta_3 < 0$)

- If $\eta_1 \geq 0$, then
  - if $\eta_2 \geq \eta_3 \left(\frac{1-\alpha\nu_\theta^-}{1-\alpha\nu_\theta^+}\right)^2$ then $\nu^* = (\nu_\theta^+,\, \nu_\theta^+,\, \nu_\theta^-)$.
  - Else $\nu^* = (\nu_\theta^+,\, \nu_2,\, \nu_3)$, where ν_2 and ν_3 are given by (2.31).
- Else ($\eta_1 < 0$) calculate ν_1 by formula (2.32).
  - If $\nu_\theta^+ < \nu_1$, then both ν_2 and ν_3 are given by (2.32).
  - Else, calculate ν_2 by formula (2.31).
    - If $\nu_\theta^+ < \nu_2$, then $\nu_1 = \nu_\theta^+$ and ν_3 is given by (2.31).
    - Else $\nu^* = (\nu_\theta^+,\, \nu_\theta^+,\, \nu_\theta^-)$.

Proof. Let
2.9, the optimal $\nu^* = (\nu_\theta^+, \nu_\theta^+, \nu_\theta^-)$.

Assume that $\eta_3 > 0$. This gives the possible cases I and III of Theorem 2.9. If we calculate ν_i, $i = 1, 2, 3$ with formula (2.34), the condition $\nu_\theta^+ < \nu_1$ is equivalent to

$$\sqrt{\eta_2} + \sqrt{\eta_3} > \sqrt{\eta_1}\left(1 + \frac{\beta\nu_\theta^+ - 1}{\beta\nu_\theta^- - 1}\right). \tag{2.66}$$

Therefore, if (2.66) is satisfied (which is case III.2 of Theorem 2.9), the optimal (ν_1, ν_2, ν_3) is given by formula (2.34). Otherwise, $\nu_1 = \nu_\theta^+$, and if we calculate ν_2, ν_3 with formula (2.33), the condition $\nu_\theta^+ < \nu_2$ is equivalent to

$$\eta_2 < \eta_3 \left(\frac{\beta\nu_\theta^- - 1}{\beta\nu_\theta^+ - 1}\right)^2. \tag{2.67}$$

This gives case III.1 of Theorem 2.9. If (2.67) is not satisfied, then $\nu^* = (\nu_\theta^+, \nu_\theta^+, \nu_\theta^-)$ is optimal, as it is asserted in case I of Theorem 2.9.

For part C of the corollary, assume that $\eta_3 > 0$, which implies the possible cases I and II of Theorem 2.9. Additionally, if $\eta_1 \geq 0$ and the inequality

$$\eta_2 < \eta_3 \left(\frac{1 - \alpha \nu_\theta^-}{1 - \alpha \nu_\theta^+} \right)^2 \tag{2.68}$$

is fulfilled, then case II.1 of Theorem 2.9 is obtained, and $\nu^* = (\nu_\theta^+, \nu_2, \nu_3)$ is optimal, where ν_2 and ν_3 are given by (2.31). If (2.68) is not satisfied, then case I of Theorem 2.9 implies that $\nu^* = (\nu_\theta^+, \nu_\theta^+, \nu_\theta^-)$ is optimal.

On the other hand, when $\eta_1 < 0$, we calculate ν_i, $i = 1, 2, 3$ with formula (2.32) and the condition $\nu_\theta^+ < \nu_1$ is equivalent to

$$\sqrt{-\eta_2} + \sqrt{-\eta_3} < \sqrt{-\eta_1} \left(1 + \frac{1 - \alpha \nu_\theta^+}{1 - \alpha \nu_\theta^-} \right). \tag{2.69}$$

If (2.69) is satisfied, then from case II.2 of Theorem 2.9 the optimal ν^* is given by formula (2.32). Otherwise, $\nu_1 = \nu_\theta^+$ and we calculate ν_2 and ν_3 with formula (2.31). The condition $\nu_\theta^+ < \nu_2$ is equivalent to

$$\eta_2 < \eta_3 \left(\frac{1 - \alpha \nu_\theta^-}{1 - \alpha \nu_\theta^+} \right)^2, \tag{2.70}$$

and if it is satisfied, then case II.2 of Theorem 2.9 implies that the optimal point is $\nu^* = (\nu_\theta^+, \nu_2, \nu_3)$, where ν_2, ν_3 are given by (2.31). Otherwise, the assumptions of part I of Theorem 2.9 are satisfied and $\nu^* = (\nu_\theta^+, \nu_\theta^+, \nu_\theta^-)$ is optimal, which concludes the proof. $\square$

Once the optimal solution $\nu^* = (\nu_1,\, \nu_2,\, \nu_3)$ for (2.30) is determined, one can easily calculate the function $g(\theta, \mathbf{N}) = \nu_1 \eta_1 + \nu_2 \eta_2 + \nu_3 \eta_3$, as well as its partial derivative over θ. In the following theorem we shall write down more explicitly regimes under which particular formulas for the derivative of the function g with respect to θ are valid.

Theorem 2.11 For $d = 3$, given $\theta \in [0, 1]$ and matrix $\mathbf{N}$ with eigenvalues $\eta_1 \geq \eta_2 \geq \eta_3$, we have

A. If $\eta_3 = 0$, then $\dfrac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \dfrac{\beta - \alpha}{(\theta\alpha + (1-\theta)\beta)^2}(\eta_1 + \eta_2)$.

B. If $\eta_3 > 0$ and additionally $\sqrt{\eta_2} + \sqrt{\eta_3} - \sqrt{\eta_1} > 0$, it holds that

$$\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \begin{cases} \dfrac{(\beta-\alpha)(\alpha+2\beta)}{\beta} \left(\dfrac{\sqrt{\eta_1} + \sqrt{\eta_2} + \sqrt{\eta_3}}{2\theta(\alpha - \beta) + \alpha + 2\beta} \right)^2, & \theta < \theta_1^B, \\ \dfrac{\beta^2 - \alpha^2}{\beta} \left(\dfrac{\sqrt{\eta_2} + \sqrt{\eta_3}}{\theta(\alpha - \beta) + \alpha + \beta} \right)^2 + \dfrac{(\beta - \alpha)\,\eta_1}{(\theta\alpha + (1-\theta)\beta)^2}, & \theta_1^B \leq \theta < \theta_2^B, \\ \dfrac{(\beta-\alpha)\eta_3}{\alpha\beta} + \dfrac{\beta - \alpha}{(\theta\alpha + (1-\theta)\beta)^2}(\eta_1 + \eta_2), & \theta \geq \theta_2^B, \end{cases}$$

where $\theta_1^B = 1 - \dfrac{\alpha(2\sqrt{\eta_1} - \sqrt{\eta_2} - \sqrt{\eta_3})}{(\beta-\alpha)(\sqrt{\eta_2} + \sqrt{\eta_3} - \sqrt{\eta_1})}$ and $\theta_2^B = 1 - \dfrac{\alpha(\sqrt{\eta_2} - \sqrt{\eta_3})}{(\beta-\alpha)\sqrt{\eta_3}}$.
If $\sqrt{\eta_2} + \sqrt{\eta_3} - \sqrt{\eta_1} \le 0$ then we omit the first case in the above formula and substitute condition $\theta_1^B \le \theta < \theta_2^B$ in the second case with the condition $\theta < \theta_2^B$.

C. If $\eta_3 < 0$ then, if η_2 and η_1 are negative as well, we have

$$\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \begin{cases} -\dfrac{(\beta-\alpha)(2\alpha+\beta)}{\alpha}\left(\dfrac{\sqrt{-\eta_1} + \sqrt{-\eta_2} + \sqrt{-\eta_3}}{2\theta(\alpha-\beta) + 3\beta}\right)^2, & \theta > \theta_1^C, \\ -\dfrac{\beta^2 - \alpha^2}{\alpha}\left(\dfrac{\sqrt{-\eta_2} + \sqrt{-\eta_3}}{\theta(\alpha-\beta) + 2\beta}\right)^2 + \dfrac{(\beta-\alpha)\,\eta_1}{(\theta\alpha + (1-\theta)\beta)^2}, & \theta_2^C < \theta \le \theta_1^C, \\ \dfrac{(\beta-\alpha)\eta_3}{\alpha\beta} + \dfrac{\beta-\alpha}{(\theta\alpha + (1-\theta)\beta)^2}(\eta_1 + \eta_2), & \theta \le \theta_2^C, \end{cases} \tag{2.71}$$

where $\theta_1^C = \dfrac{\beta(\sqrt{-\eta_2} + \sqrt{-\eta_3} - 2\sqrt{-\eta_1})}{(\beta-\alpha)(\sqrt{-\eta_2} + \sqrt{-\eta_3} - \sqrt{-\eta_1})}$ and $\theta_2^C = \dfrac{\beta(\sqrt{-\eta_3} - \sqrt{-\eta_2})}{(\beta-\alpha)\sqrt{-\eta_3}}$.
If $\eta_2 < 0$ and $\eta_1 \ge 0$ then θ_1^C is not defined and we can express $\frac{\partial g}{\partial \theta}(\theta, \mathbf{N})$ by the second and the third term in (2.71), omitting the assumption $\theta \le \theta_1^C$ in the second case.
If $\eta_2 \ge 0$ then neither θ_1^C nor θ_2^C are defined and $\frac{\partial g}{\partial \theta}$ is given by the formula in the third case of (2.71), for any $\theta \in [0, 1]$.

Proof. Let us first remark that $\theta_1^B \le \theta_2^B \le 1$ and $0 \le \theta_2^C \le \theta_1^C$. If $\eta_3 = 0$, then case A of Corollary 2.10 implies that $(\nu_\theta^+, \nu_\theta^+, \nu_\theta^-)$ is optimal and $g(\theta, \mathbf{N}) = \nu_\theta^+(\eta_1 + \eta_2)$. By taking the derivative of the function g with respect to θ, case A of the theorem is obtained.

Let us assume that $\eta_3 > 0$. If the optimal (ν_1, ν_2, ν_3) is given by formula (2.34), then by taking the derivative of the function

$$g(\theta, \mathbf{N}) = \eta_1\nu_1 + \eta_2\nu_2 + \eta_3\nu_3$$

with respect to θ gives the first term in case B. The condition for this case, $\nu_\theta^+ < \nu_1$, is equivalent to $\theta < \theta_1^B$. If $\nu_\theta^+ = \nu_1$ (or equivalently, $\theta \ge \theta_1^B$) and ν_2 and ν_3 are given by (2.33), the second term for $\frac{\partial g}{\partial \theta}$ in case B is obtained. This case occurs if $\nu_\theta^+ < \nu_2$ which is equivalent to $\theta < \theta_2^B$. Finally, the last term in case B is easily reconstructed since in this case $(\nu_\theta^+, \nu_\theta^+, \nu_\theta^-)$ is optimal for the minimization problem in the definition of function g.

If $\eta_1 < 0$ (which implies that $\eta_2, \eta_3 < 0$) and the optimal (ν_1, ν_2, ν_3) is given by formula (2.32), then from

$$g(\theta, \mathbf{N}) = \eta_1\nu_1 + \eta_2\nu_2 + \eta_3\nu_3$$

an easy calculation gives us the formula which corresponds to the first term of the function $\frac{\partial g}{\partial \theta}$ in the case C. Here, the condition $\nu_\theta^+ < \nu_1$ is equivalent to $\theta > \theta_1^C$. If $\nu_1 = \nu_\theta^+$ (or equivalently, $\theta \le \theta_1^C$) and ν_2 and ν_3 are given by (2.31), then one gets the second formula in case C. This occurs if $\nu_\theta^+ < \nu_2$ or equivalently $\theta > \theta_2^C$. The last term in case C follows easily, since in this case $(\nu_\theta^+, \nu_\theta^+, \nu_\theta^-)$ is optimal for the minimization problem in the definition of function g. $\square$

It is important to notice that the function $\theta \mapsto g_\alpha(\mathbf{x}, \mathsf{u}^k(\mathbf{x})) - g_\beta(\mathbf{x}, \mathsf{u}^k(\mathbf{x})) + l + \frac{\partial g}{\partial \theta}(\theta, \mathbf{N}^k(\mathbf{x}))$ is continuous in the three-dimensional case, due to the continuity of the function $\theta \mapsto \frac{\partial g}{\partial \theta}$, but not necessarily monotone as it was in the two-dimensional case. The possible lack of monotonicity can occur in case C, when $\eta_1 \geq 0$, $\eta_2 < 0$, and for some choices of α, β, η_1, η_2, η_3. In this case, one can get two possible zeros of this function on $[0, 1]$, and then we simply take the smaller one for the next iteration of θ. In all other cases, function (2.27) is monotone and its zero is explicitly calculated by solving a quadratic (or quartic) equation.

2.2.2 Numerical examples

In this section, we apply Algorithm 2.7 on several optimal design problems. The state and adjoint equations are solved by the finite element method from the deal.II library (Bangerth et al. (2000)) using Lagrange elements on a fine mesh, while a design $(\theta, \mathbf{A})$ is discretized on a (possibly different) mesh (Casado-Díaz et al. (2011)), by piecewise constant elements. The Lagrange multiplier l is recalculated at each step in a way that θ^{k+1} satisfies the volume constraint, which is done quite effectively by the bisection method. All problems are treated for various volume fractions $\eta := \frac{q_\alpha}{|\Omega|}$ of the first phase (with conductivity α). For the initial design we take constant $\theta^0 = \eta$, while $\mathbf{A}^0$ is taken to be a simple laminate ($\mathbf{A}^0 = \operatorname{diag}(\lambda_\theta^-, \lambda_\theta^+)$ if $d = 2$ or $\mathbf{A}^0 = \operatorname{diag}(\lambda_\theta^-, \lambda_\theta^+, \lambda_\theta^+)$ if $d = 3$). In all examples we calculate 20 iterations of Algorithm 2.7, but it appears that optimal design is well approximated by the first several iterations.

Example 2.1 (Two-state problem on a ball.) In the first example we consider a two-dimensional problem of weighted energy minimization

$$J(\theta, \mathbf{A}) = 2\int_\Omega f_1 u_1 \, d\mathbf{x} + \int_\Omega f_2 u_2 \, d\mathbf{x} \longrightarrow \min,$$

where $\Omega \subseteq \mathbf{R}^2$ is a ball $B(\mathbf{0}, 2)$, $\alpha = 1$, $\beta = 2$, while u_1 and u_2 are state functions for

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u_i) = f_i & \text{in } \Omega \\ u_i \in \mathrm{H}_0^1(\Omega) \end{cases}, \quad i = 1, 2, \tag{2.72}$$

where we take $f_1 = \chi_{B(\mathbf{0},1)}$ and $f_2 \equiv 1$ for right-hand sides. This problem is explicitly solved by Burazin & Vrdoljak (2018) so we can compare our numerical solution to the exact one. The comparison is done with respect to mesh refinement: the original triangulation of the domain is refined up to 8 times, where each refinement introduces a four times finer mesh (Bangerth et al. (2000)). The L^1 error between the numerical and exact solutions

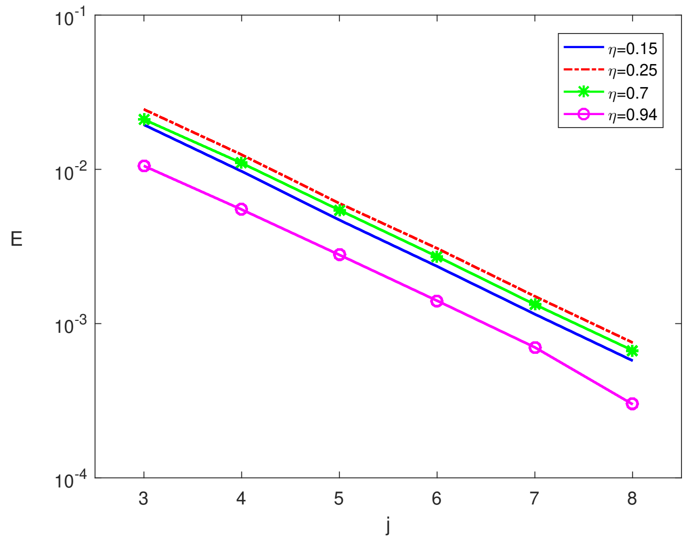


Figure 2.1: L^1 norm E of the difference between the numerical and exact solutions with respect to mesh refinement j (each refinement introduces a four times finer mesh) for various choices of volume fractions η of the first phase (Example 2.1).

is presented on Figure 2.1 for various choices of η, and it can be seen that the numerical solution gives a good approximation to the exact one.

For $\eta = 0.25$, the numerical solution is presented in Figure 2.2. Let us recall that $\theta = 0$ corresponds to the material with conductivity β, $\theta = 1$ corresponds to the material with conductivity α, while $\theta \in \langle 0, 1\rangle$ corresponds to a fine mixture of the original phases. Convergence history is presented in Figure 2.3.

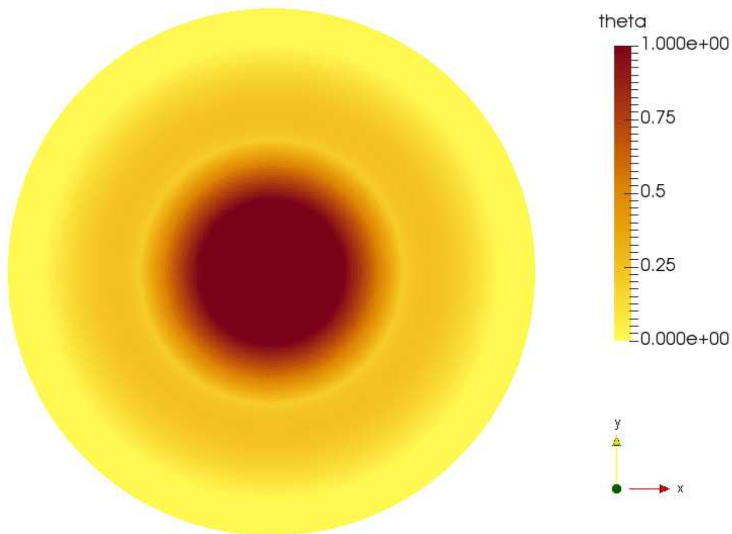
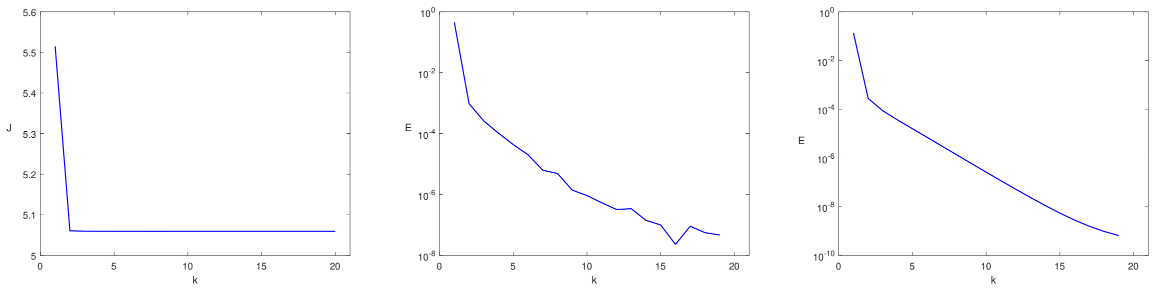


Figure 2.2: Numerical solution for the optimal design problem presented in Example 2.1.



(a) Cost functional J in terms of the iteration number k. **(b)** $E = |J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ in terms of the iteration number k. **(c)** $E = \|\theta^k - \theta^{k+1}\|_{L^2}^2$ in terms of the iteration number k.

Figure 2.3: Convergence history for the optimal design problem presented in Example 2.1.

Example 2.2 (Single state problem on an annulus.) Let us now consider the energy minimization problem

$$J(\theta, \mathbf{A}) = \int_\Omega f u \, d\mathbf{x} \longrightarrow \min,$$

within an annulus $B(\mathbf{0}; 1, 2) \subseteq \mathbf{R}^2$, with inner radius 1 and outer radius 2 and the state equation

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u) = 1 & \text{in } \Omega \\ u \in \mathrm{H}_0^1(\Omega)\,. \end{cases} \tag{2.73}$$

The exact solution for this example is calculated in Burazin (2018), which again allows us to compare our numerical solution to the exact one. The L^1 error between the numerical and exact solutions is given in Figure 2.4 for various $0 < \eta < 1$ and it is a decreasing function with respect to mesh refinement.

Optimal distribution with 50% of the first material is shown in Figure 2.5, while convergence histories of the cost functional and the approximation error are illustrated in Figure 2.6.

We can conclude for both examples that the optimality criteria method proposed in Section 2.2 gives a good approximation of the exact solution.

Example 2.3 (Two-state problem on a cube.) The third example is the three-dimensional energy minimization problem

$$J(\theta, \mathbf{A}) = \int_\Omega (f_1 u_1 + f_2 u_2) \, d\mathbf{x} \longrightarrow \min,$$

with $\alpha = 1$, $\beta = 2$ and two state equations

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u_i) = f_i & \text{in } \Omega \\ u_i \in \mathrm{H}_0^1(\Omega) \end{cases}, \quad i = 1, 2. \tag{2.74}$$

We take a cube $\Omega = [-1, 1]^3$ as the domain and set function f_1 to be zero on the upper

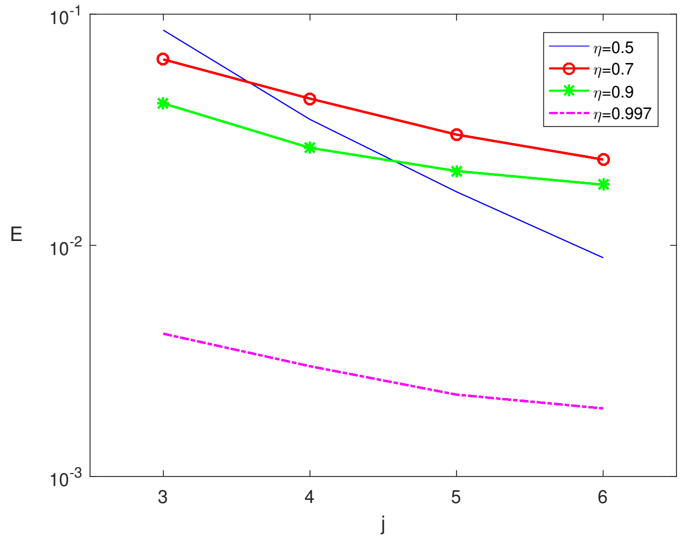


Figure 2.4: L^1 norm E of the difference between the numerical and exact solutions with respect to mesh refinement j (each refinement introduces a four times finer mesh) for various choices of volume fraction η of the first phase (Example 2.2).

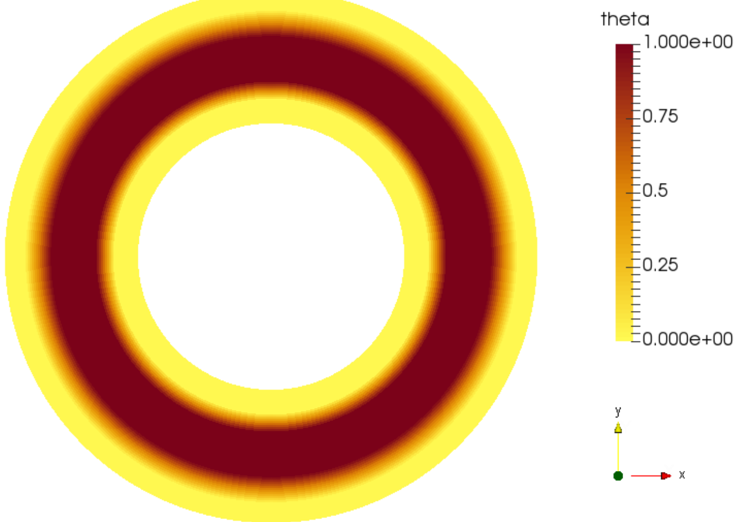
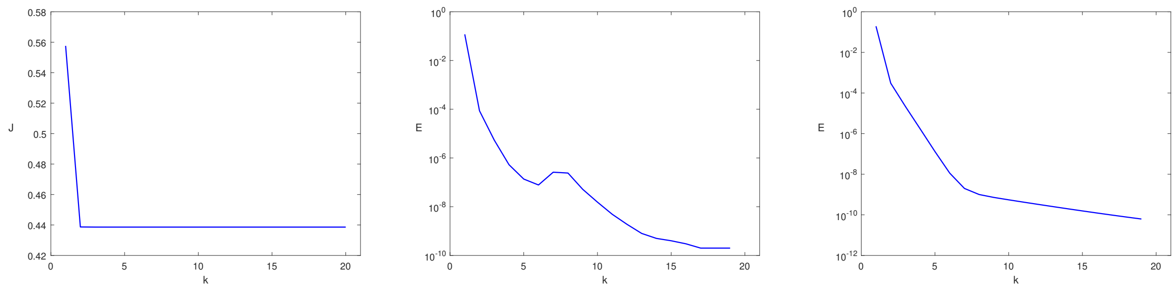


Figure 2.5: Numerical solution for the optimal design problem presented in Example 2.2.



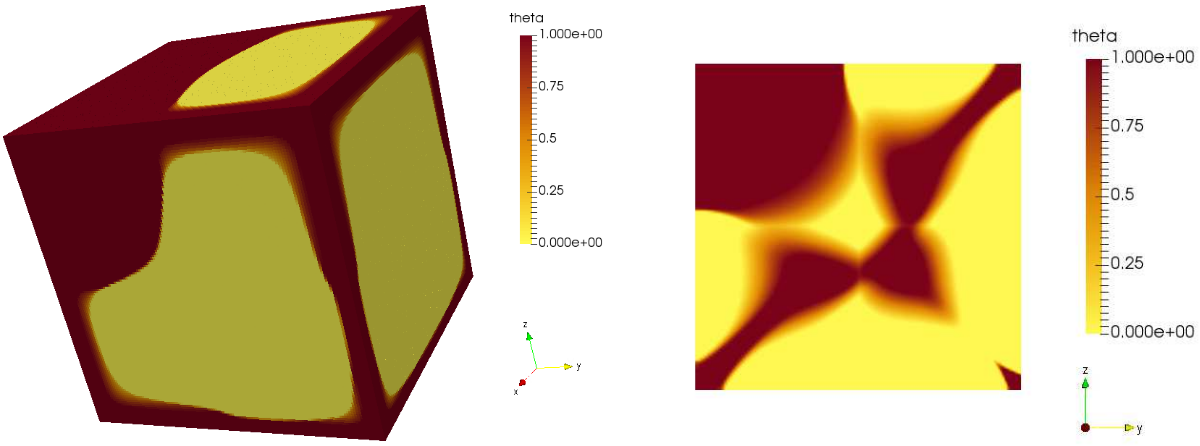
(a) Cost functional J in terms of the iteration number k.

(b) $E = |J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ in terms of the iteration number k.

(c) $E = \|\theta^k - \theta^{k+1}\|_{L^2}^2$ in terms of the iteration number k.

Figure 2.6: Convergence history for the optimal design problem presented in Example 2.2.

half ($z > 0$) and 10 on the lower half of the cube, while function f_2 to be zero on the left half ($y < 0$) and 10 on the right half of the cube. The optimal design of the 20-th iteration of the Algorithm 2.7 with volume fraction $\eta = 0.5$ of the first material is shown in Figure 2.7a.



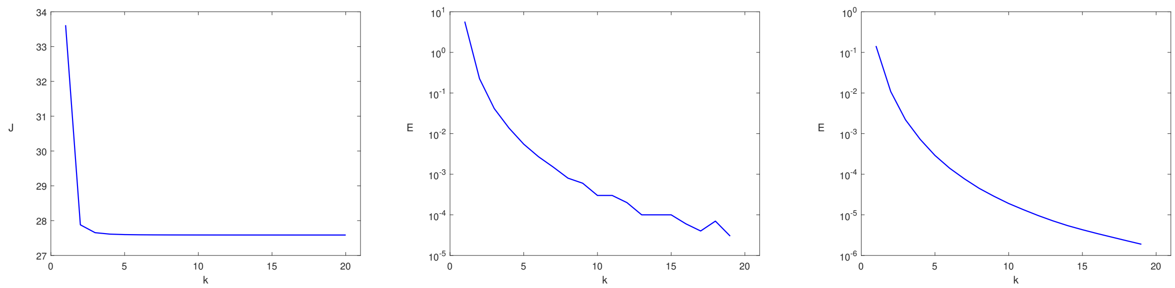
(a) Outer look.

(b) Intersection of the cube with the plane $x = 0$.

Figure 2.7: Numerical solution for the optimal design problem presented in Example 2.3.

Material with greater conductivity is placed at the center of the cube and on the sides, which can be seen in Figure 2.7b. Most of the upper left part of the cube is occupied by the material with smaller conductivity, which is expected because there is no external source on this part of the domain. The convergence history of the cost functional and the residual is shown in Figure 2.8.

Example 2.4 (Non self-adjoint problem on a cube.) Let us now consider a non self-adjoint



(a) Cost functional J in terms of the iteration number k.

(b) $E = |J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ in terms of the iteration number k.

(c) $E = \|\theta^k - \theta^{k+1}\|^2_{L^2}$ in terms of the iteration number k.

Figure 2.8: Convergence history for the optimal design problem presented in Example 2.3.

two-state minimization problem, where the cost functional is given by

$$J(\theta, \mathbf{A}) = \int_\Omega (u_1^2 + u_2^2)\, d\mathbf{x}.$$

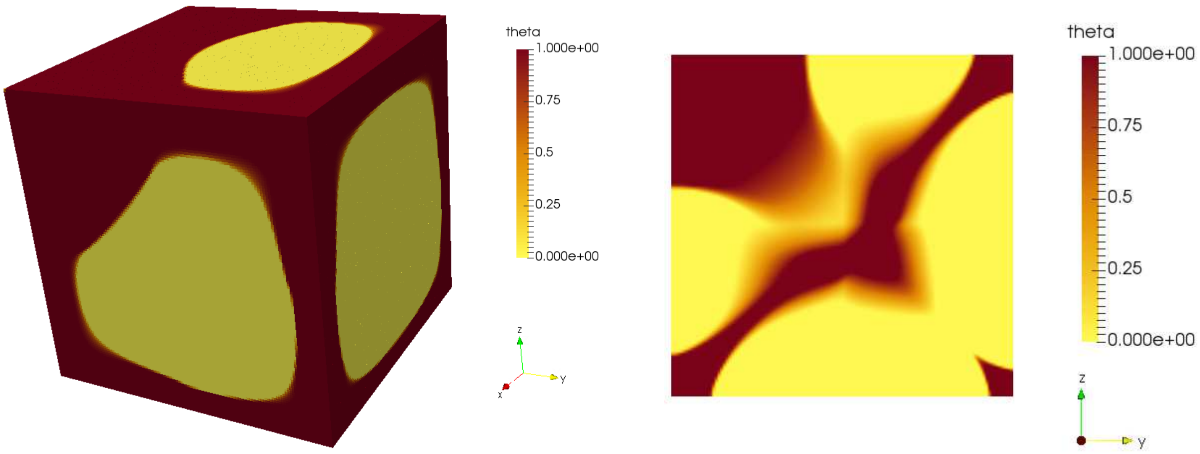
We take state equations (2.74) and domain $\Omega = [-1, 1]^3$, with f_1 and f_2 being as in Example 2.3. In this case, the adjoint equations are given by

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla p_i) = 2u_i & \text{in } \Omega \\ p_i \in \mathrm{H}^1_0(\Omega) \end{cases}, \quad i = 1, 2. \tag{2.75}$$

A numerical solution to this optimal design problem with $\alpha = 1$, $\beta = 2$ and volume fraction $\eta = 0.5$ of the first material is presented in Figure 2.9a, while the intersection of the domain with the plane $x = 0$ is shown in Figure 2.9b. Convergence history is shown in Figure 2.10.

For the last two examples exact solutions are not known, and therefore, we cannot make the comparison like we could for the first two examples.

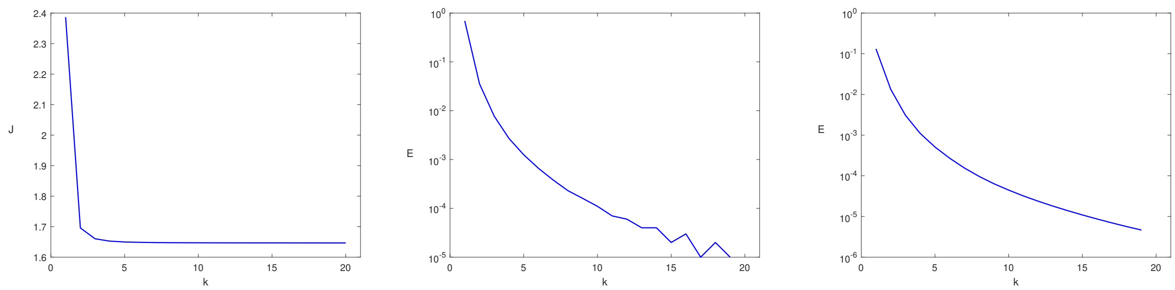
As was mentioned at the beginning of the chapter, the first variant of the optimality criteria method for multiple state optimal design problems (2.9) is presented in Vrdoljak (2010), but it does not converge for examples presented in this subsection. On the other hand, that variant behaves well for the question of maximization of the same functionals instead of minimization. This kind of behaviour is expected for a class of self-adjoint problems, since already single state self-adjoint problems exhibit a similar effect (Murat & Tartar (1985), Allaire (2002)).



(a) Outer look.

(b) Intersection of the cube with the plane $x = 0$.

Figure 2.9: Numerical solution for the optimal design problem presented in Example 2.4.



(a) Cost functional J in terms of the iteration number k.

(b) $E = |J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ in terms of the iteration number k.

(c) $E = \|\theta^k - \theta^{k+1}\|_{L^2}^2$ in terms of the iteration number k.

Figure 2.10: Convergence history for the optimal design problem presented in Example 2.4.

2.2.3 Convergence proof

Algorithm 2.7 appears to give a minimizing sequence of designs in the case of minimization of a conic sum of energies. Thus, in the sequel we restrict ourselves to this particular case. Namely, we take $g_\alpha = g_\beta = \sum_{i=1}^m \mu_i f_i u_i$, for some $\mu_i > 0$, which imposes the following simplifications: for $i = 1, \dots, m$, we have $p_i^* = \mu_i u_i^*$, which implies $\boldsymbol{\tau}_i^* = \mu_i \boldsymbol{\sigma}_i^*$ and $\mathbf{N}^* = \sum_{i=1}^m \mu_i (\boldsymbol{\sigma}_i^* \otimes \boldsymbol{\sigma}_i^*)$.

Remark 2.7. Note that in this case of minimizing the conic sum of energies, case C in Theorem 2.11 never occurs, since the matrix $\mathbf{N}^*$ is positive semidefinite. Therefore, in each step of Algorithm 2.7 the function $\theta \mapsto R^k(\theta, \mathbf{x})$ is monontone and its zero (if exist) is explicitly calculated for a.e. $\mathbf{x} \in \Omega$.

The proof of convergence for a variant of the optimality criteria method for single state self-adjoint minimization problems was given in Allaire (2002) (the original proof of convergence is due to Toader (1997)). It relies on the fact that the optimal design can be found among simple laminates (Murat & Tartar (1985)). In general, this is not the truth for multiple state problems (Allaire (2002), Vrdoljak (2010), see also Theorem 2.9) anymore, and thus the proof cannot be generalized to this case. However, a recent work of Burazin and Vrdoljak (Burazin & Vrdoljak (2018)) on minimizing the conic sum of energies shows that the optimal design can be found among simple laminates in two specific cases: when the number of states is less then the space dimension or in the spherically symmetric case. This enables us to derive the convergence proof of Algorithm 2.7, but first we need some additional information regarding the output of the algorithm in these two cases.

Case $m < d$

If the number of states is less then the space dimension, it was recently shown (Burazin & Vrdoljak (2018)) that even for the multiple state problems the optimal $\mathbf{A}^*$ can be achieved among simple laminates, with the lamination direction orthogonal to the corresponding fluxes $\boldsymbol{\sigma}_1^*, \dots, \boldsymbol{\sigma}_m^*$. We are going to prove that in this case Algorithm 2.7 in each iteration gives $\mathbf{A}^{k+1}$ which is also a simple laminate with the lamination direction orthogonal to $\boldsymbol{\sigma}_1^k, \dots, \boldsymbol{\sigma}_m^k$. We shall need the following lemma.

Lemma 2.12 Let $\mu_i > 0$ and $\boldsymbol{\sigma}_i \in \mathbf{R}^d$, $i = 1, \dots m$. If $\mathbf{N} = \sum_{i=1}^m \mu_i(\boldsymbol{\sigma}_i \otimes \boldsymbol{\sigma}_i)$ then $\mathcal{R}(\mathbf{N}) = \mathrm{Span}(\{\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m\})$.

Proof. The inclusion $\mathcal{R}(\mathbf{N}) \subseteq \mathrm{Span}(\{\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m\})$ is trivial. Conversely, since the matrix $\mathbf{N}$, whose eigenvalues we denote by $\eta_1 \geq \eta_2 \geq \cdots \geq \eta_d \geq 0$, is a sum of m matrices of rank 1, then $r := \mathrm{rank}(\mathbf{N}) \leq m$ and $\eta_{r+1} = \cdots = \eta_d = 0$. Moreover, it is a real, symmetric matrix, therefore diagonalizable, and its eigenvectors $\mathsf{e}_1, \dots, \mathsf{e}_d$ constitite an orthonormal basis for $\mathbf{R}^d$. It follows

$$\mathbf{N} = \sum_{i=1}^r \eta_i(\mathsf{e}_i \otimes \mathsf{e}_i),$$

and since eigenvectors of matrix $\mathbf{N}$ that correspond to the positive eigenvalues form a basis for $\mathcal{R}(\mathbf{N})$, it is sufficient to prove that every $\boldsymbol{\sigma}_j$, $j = 1, \dots, m$, is a linear combination of $\mathsf{e}_1, \dots, \mathsf{e}_r$. If this is not the case, then there exists $j \in \{1, \dots, m\}$ and $k \in \{r+1, \dots, d\}$ such that $\boldsymbol{\sigma}_j \cdot \mathsf{e}_k \neq 0$. However, this would imply

$$0 = \mathbf{N}\mathsf{e}_k \cdot \mathsf{e}_k = \left(\sum_{i=1}^m \mu_i(\boldsymbol{\sigma}_i \cdot \mathsf{e}_k)\boldsymbol{\sigma}_i\right) \cdot \mathsf{e}_k = \sum_{i=1}^m \mu_i(\boldsymbol{\sigma}_i \cdot \mathsf{e}_k)^2,$$

which is a contradiction. Therefore, $\mathcal{R}(\mathbf{N}) = \mathrm{Span}(\{\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m\})$. □

From Lemma 2.12 we conclude that in each step of Algorithm 2.7, if the number of states is less then the space dimension, then $\mathcal{R}(\mathbf{N}^k) = \mathrm{Span}(\{\boldsymbol{\sigma}_1^k, \dots, \boldsymbol{\sigma}_m^k\})$. This also implies that the matrix $\mathbf{N}^k$ has $d - r > 0$ eigenvalues equal to zero, for r being the dimension of $\mathrm{Span}(\{\boldsymbol{\sigma}_1^k, \dots, \boldsymbol{\sigma}_m^k\})$. Thus, by Remark 2.4, $\mathbf{A}^{k+1}$ from step 4 of Algorithm 2.7 is a simple laminate with the lamination direction from $\ker \mathbf{N}^k$, and thus orthogonal to $\mathcal{R}(\mathbf{N}^k)$, which proves the following theorem.

Theorem 2.13 If $m < d$, then Algorithm 2.7 in step 4 gives $\mathbf{A}^{k+1}$ as a simple laminate with the lamination direction orthogonal to each $\boldsymbol{\sigma}_i^k$, $i = 1, \dots, m$.

□

Remark 2.8. Notice that the fourth step of Algorithm 2.7 is not executed only when $\mathbf{N}^k(\mathbf{x}) = \mathbf{0}$, i. e. when all $\boldsymbol{\sigma}_i^k$ equal $\mathsf{0}$. In this case we take $\mathbf{A}^{k+1}(\mathbf{x}) = \mathbf{A}^k(\mathbf{x})$. If we want to ensure that our algorithm gives also simple laminates as output in this case, we can simply take the initial $\mathbf{A}^0$ to be a simple laminate. However, for the proof of convergence this is not necessary, as we shall only use that $\mathbf{A}^{k+1}\boldsymbol{\sigma}_i^k = \lambda^+_{\theta^{k+1}}\boldsymbol{\sigma}_i^k$, which is here trivially satisfied.

Spherically symmetric case

Let us now consider a case of spherical symmetry: in this subsection, we assume that the domain Ω is spherically symmetric and that the right-hand sides $f_i \in \mathrm{L}^2(\Omega)$ of state equations (1.1) are radial functions, i.e. $f_i = f_i(r)$, for $r = |\mathbf{x}|$. In other words, $f_i(\mathbf{Q}\mathbf{x}) = f_i(\mathbf{x})$, $i = 1, \dots, m$ for all orthogonal matrices $\mathbf{Q}$, i.e. matrices belonging to the symmetry group of Ω

$$O(d) = \{\mathbf{Q} \in \mathrm{M}_d(\mathbf{R}) : \mathbf{Q}^\top\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\top = \mathbf{I}\}.$$

In Burazin & Vrdoljak (2018) it was proved that, in the spherically symmetric case, there is an optimal θ^* which is radial, and there is an optimal $\mathbf{A}^*$, which is a simple laminate with the lamination direction orthogonal to the radial vector e_r. Moreover, since u_i^* are radial functions (Burazin & Vrdoljak 2018, Lemma 4), ∇u_i^* and, consequently $\boldsymbol{\sigma}_i^*$, are collinear with e_r, for $i = 1, \dots, m$. To prove that this also happens in every iteration of Algorithm 2.7, we shall need the following lemma.

Lemma 2.14 In the spherically symmetric case, if $\mathbf{A}(\mathbf{x})\mathsf{e}_r = a(|\mathbf{x}|)\mathsf{e}_r$, for some bounded and coercive function a, then the solution of the equation

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u) = f & \text{in } \Omega \\ u \in \mathrm{H}_0^1(\Omega) \end{cases} \tag{2.76}$$

is a radial function.

Proof. Let us denote by $(r, \boldsymbol{\varphi}) = (r, \varphi_1, \varphi_2, \dots, \varphi_{d-1})$ the spherical coordinates in $\mathbf{R}^d$ in which Ω can be represented as $r \in \langle R_1, R_2\rangle$, $\boldsymbol{\varphi} \in S := [0, \pi] \times \cdots \times [0, \pi] \times [0, 2\pi]$, for some $0 < R_1 < R_2$. In spherical coordinates equation (2.76) becomes

$$\begin{cases} -(r^{d-1}a(r)\tilde{u}'(r))' = r^{d-1}f(r) & \text{in } \langle R_1,\, R_2\rangle \\ \tilde{u} \in \mathrm{H}_0^1(\langle R_1,\, R_2\rangle), \end{cases} \tag{2.77}$$

under the assumption that $u(\mathbf{x}) = \tilde{u}(|\mathbf{x}|)$. It can easily be seen that a radial solution of corresponding equation is spherical coordinates (2.77) satisfies equation (2.76) in the sense of distributions, and therefore, it is solution for (2.76). □

Theorem 2.15 In the spherically symmetric case, if the initial θ^0 is a radial function and $\mathbf{A}^0 \in \mathcal{K}(\theta^0)$ is a simple laminate with the lamination direction orthogonal to the radial vector e_r, Algorithm 2.7 in each iteration gives a radial θ^{k+1} and $\mathbf{A}^{k+1}$ as a simple laminate with the lamination direction orthogonal to every $\boldsymbol{\sigma}_i^k$, $i = 1, \ldots, m$.

Proof. By induction, let $(\theta^k, \mathbf{A}^k)$ be a design obtained in k-th iteration of the algorithm, where θ^k is a radial function and $\mathbf{A}^k$ a simple laminate with the lamination direction orthogonal to $\boldsymbol{\sigma}_i^{k-1}$, $i = 1, \ldots, m$. Since f_i are radial functions and $\mathbf{A}^k\mathsf{e}_r = \lambda^+_{\theta^k(|\mathbf{x}|)}\mathsf{e}_r$, by Lemma 2.14 u_i^k are also radial functions for $i = 1, \ldots, m$. Therefore, $\boldsymbol{\sigma}_i^k$ are collinear with the radial vector, i.e. $\boldsymbol{\sigma}_i^k = \varphi_i^k(r)\mathsf{e}_r$, for some functions φ_i^k, $i = 1, \ldots, m$, implying that

$$\mathbf{N}^k = \sum_{i=1}^m \mu_i \boldsymbol{\sigma}_i^k \otimes \boldsymbol{\sigma}_i^k = \sum_{i=1}^m \mu_i (\varphi_i^k(r))^2 \mathsf{e}_r \otimes \mathsf{e}_r.$$

From the above equality, we conclude that the matrix $\mathbf{N}^k$ has one positive and radial eigenvalue $\eta_1^k = \sum_{i=1}^m \mu_i(\varphi_i^k(r))^2 = \sum_{i=1}^m \mu_i |\boldsymbol{\sigma}_i^k|^2$, that corresponds to the eigenvector e_r, while all others eigenvalues are equal to zero. Since θ^{k+1} depends on $\mathbf{N}^k$, which is a function of $r = |\mathbf{x}|$, it follows that it is a radial function. Moreover, from Remark 2.4, it is clear that the next $\mathbf{A}^{k+1}$ is a simple laminate with the lamination direction orthogonal to e_r, and the proof follows from the collinearity of every $\boldsymbol{\sigma}_i^k$ with e_r.

□

We shall prove the convergence of Algorithm 2.7 for the multiple state minimization of a conic sum of energies in two cases: the spherically symmetric case, and the case when the number of states is less then the space dimension. The partial derivative of the function $g(\theta, \mathbf{N}) = \min\limits_{\mathbf{A}\in\mathcal{K}(\theta)} (\mathbf{A}^{-1} : \mathbf{N})$ is given in theorems 2.8 and 2.11 (for two and three space dimensions), but in these two special cases, the partial derivative of g takes a simpler form, which will be seen from the following lemma.

Lemma 2.16 Let $\mu_i > 0$, $\boldsymbol{\sigma}_i \in \mathbf{R}^d$, $i = 1, \ldots, m$, and $\mathbf{N} = \sum_{i=1}^m \mu_i(\boldsymbol{\sigma}_i \otimes \boldsymbol{\sigma}_i)$. If $\mathrm{rank}(\mathbf{N}) < d$, then

$$\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \frac{(\beta - \alpha)}{(\lambda_\theta^+)^2} \sum_{i=1}^m \mu_i |\boldsymbol{\sigma}_i|^2. \tag{2.78}$$

Proof. Let $\eta_1 \geq \eta_2 \geq \dots \geq \eta_d \geq 0$ be eigenvaules of the matrix $\mathbf{N}$. Since $r = \mathrm{rank}(\mathbf{N}) < d$, then $d - r$ of these eigenvalues are equal to zero, and

$$g(\theta, \mathbf{N}) = \min_{\mathbf{A} \in \mathcal{K}(\theta)} (\mathbf{A}^{-1} : \mathbf{N}) = \frac{1}{\lambda_\theta^+} \sum_{i=1}^{r} \eta_i.$$

Therefore,

$$\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}) = \frac{(\beta - \alpha)}{(\lambda_\theta^+)^2} \sum_{i=1}^{r} \eta_i.$$

Using the trace theorem, we get

$$\sum_{i=1}^{r} \eta_i = \mathrm{tr}(\mathbf{N}) = \mathrm{tr}\left(\sum_{i=1}^{m} \mu_i (\boldsymbol{\sigma}_i \otimes \boldsymbol{\sigma}_i) \right) = \sum_{i=1}^{m} \mu_i |\boldsymbol{\sigma}_i|^2,$$

and the proof is complete. □

As it was shown in theorems 2.13 and 2.15, for the spherically symmetric case and the case when the number of states is less then the space dimension, in each iteration of Algorithm 2.7, $\mathbf{A}^{k+1}$ is a simple laminate with the lamination direction orthogonal to every $\boldsymbol{\sigma}_1^k, \boldsymbol{\sigma}_2^k, \dots, \boldsymbol{\sigma}_m^k$. The previous lemma implies that in each iteration of the algorithm we have $\frac{\partial g}{\partial \theta}(\theta, \mathbf{N}^k) = \frac{(\beta - \alpha)}{(\lambda_\theta^+)^2} \sum_{i=1}^{m} \mu_i |\boldsymbol{\sigma}_i^k|^2$, which enables us to rewrite the algorithm in the following way:

Algorithm 2.17 Initialize $(\theta^0, \mathbf{A}^0) \in \mathcal{A}$ (in the spherically symmetric case take a radial θ^0 and $\mathbf{A}^0$ as a simple laminate with the lamination direction orthogonal to e_r). For $k \geq 0$:

(1) Calculate the solution u_i^k, $i = 1, \dots, m$ of

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^k \nabla u_i^k) = f_i & \text{in } \Omega \\ u_i^k \in \mathrm{H}_0^1(\Omega) \end{cases}$$

and define $\boldsymbol{\sigma}_i^k := \mathbf{A}^k \nabla u_i^k$ and $\mathbf{N}^k := \sum_{i=1}^{m} \mu_i (\boldsymbol{\sigma}_i^k \otimes \boldsymbol{\sigma}_i^k)$.

If $\mathbf{N}^k(\mathbf{x}) = \mathbf{0}$, for $\mathbf{x} \in \Omega$, leave the old data for the next iteration of $\theta^{k+1}(\mathbf{x})$ and $\mathbf{A}^{k+1}(\mathbf{x})$. Else, do step 2.

(2) For $\mathbf{x} \in \Omega$ take $\theta^{k+1}(\mathbf{x})$ as a zero of the function

$$\theta \mapsto R^k(\theta, \mathbf{x}) := l + \frac{(\beta - \alpha)}{(\lambda_\theta^+)^2} \sum_{i=1}^{m} \mu_i |\boldsymbol{\sigma}_i^k|^2\,, \tag{2.79}$$

and if a zero doesn't exist, take 0 (or 1) if the function is positive (or negative) on $[0, 1]$.

Diagonalize the matrix $\mathbf{N}^k(\mathbf{x})$ and let $\mathbf{A}^{k+1}(\mathbf{x})$ be a simple laminate with the lamination direction orthogonal to each $\boldsymbol{\sigma}_i^k(\mathbf{x})$, $i = 1, \dots, m$, i. e. from $\ker \mathbf{N}^k(\mathbf{x})$.

Remark 2.9. In the case of minimizing a conic sum of energies, the objective functional J in the relaxed problem (2.9) can be written via dual formulation. For $i \in \{1, \ldots, m\}$ fixed, a weak solution of state equation

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}\nabla u_i) = f_i & \text{in } \Omega \\ u_i \in \mathrm{H}_0^1(\Omega)\,, \end{cases} \tag{2.80}$$

for $\mathbf{A} \in \mathrm{L}^\infty(\Omega; \mathcal{M}_{\alpha,\beta})$ and $f_i \in \mathrm{H}^{-1}$ is the unique minimum of the quadratic functional

$$\frac{1}{2}\int_\Omega \mathbf{A}\nabla u_i \cdot \nabla u_i \, d\mathbf{x} - \int_\Omega f_i u_i \, d\mathbf{x}$$

on $\mathrm{H}_0^1(\Omega)$. From the assumption $\mathbf{A}\boldsymbol{\xi} \cdot \boldsymbol{\xi} \geq \alpha|\boldsymbol{\xi}|^2$, $\boldsymbol{\xi} \in \mathbf{R}^d$, by taking $\boldsymbol{\xi} = \nabla v_i - \mathbf{A}^{-1}\boldsymbol{\tau}_i$, for $v_i \in \mathrm{H}_0^1(\Omega)$, $\boldsymbol{\tau}_i \in \mathrm{L}^2(\Omega; \mathbf{R}^d)$, the monotonicity of the integral implies

$$\int_\Omega (\mathbf{A}\nabla v_i - \boldsymbol{\tau}_i) \cdot (\nabla v_i - \mathbf{A}^{-1}\boldsymbol{\tau}_i) \, d\mathbf{x} \geq 0, \quad \text{a.e. } \mathbf{x} \in \Omega, \tag{2.81}$$

which is equivalent to

$$-\int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_i \, d\mathbf{x} \leq \int_\Omega (\mathbf{A}\nabla v_i \cdot \nabla v_i - 2\boldsymbol{\tau}_i \cdot \nabla v_i) \, d\mathbf{x}, \quad \text{a.e. } \mathbf{x} \in \Omega.$$

Equality in the above inequality is achieved if and only of $\nabla v_i - \mathbf{A}^{-1}\boldsymbol{\tau}_i = 0$ (since $\mathbf{A}$ is positive semidefinite), which happens if and only if $\boldsymbol{\tau}_i = \mathbf{A}\nabla v_i$. By taking the maximum in the last inequality over set of all $\boldsymbol{\tau}_i \in \mathrm{L}^2(\Omega; \mathbf{R}^d)$ such that $-\mathsf{div}\,\boldsymbol{\tau}_i = f_i$ in Ω, we have

$$-\min_{\substack{\boldsymbol{\tau}_i \in \mathrm{L}^2(\Omega;\mathbf{R}^\mathrm{d}) \\ -\mathrm{div}\boldsymbol{\tau}_i = f_i \text{ in } \Omega}} \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_i = \int_\Omega (\mathbf{A}\nabla v_i \cdot \nabla v_i - 2f_i v_i) \, d\mathbf{x}, \quad v_i \in \mathrm{H}_0^1(\Omega).$$

In particular, for the solution u_i of (2.80), it follows

$$\begin{aligned} -\int_\Omega f_i u_i \, d\mathbf{x} &= \int_\Omega (\mathbf{A}\nabla u_i \cdot \nabla u_i - f_i u_i) \, d\mathbf{x} - \int_\Omega f_i u_i \, d\mathbf{x} = \\ &= \min_{v_i \in \mathrm{H}_0^1(\Omega)} \int_\Omega (\mathbf{A}\nabla v_i \cdot \nabla v_i - 2f_i v_i) \, d\mathbf{x} = \\ &= -\min_{\substack{\boldsymbol{\tau}_i \in \mathrm{L}^2(\Omega;\mathbf{R}^\mathrm{d}) \\ -\mathrm{div}\boldsymbol{\tau}_i = f_i \text{ in } \Omega}} \mathbf{A}^{-1}\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_i \, d\mathbf{x}. \end{aligned}$$

Therefore, the functional in (2.9) becomes

$$J(\theta, \mathbf{A}) = \sum_{i=1}^m \mu_i \int_\Omega f_i u_i \, d\mathbf{x} + l \int_\Omega \theta \, d\mathbf{x} = \min_{\substack{\boldsymbol{\tau}_i \in \mathrm{L}^2(\Omega;\mathbf{R}^d) \\ -\mathrm{div}\boldsymbol{\tau}_i = f_i}} \sum_{i=1}^m \mu_i \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau}_i \cdot \boldsymbol{\tau}_i \, d\mathbf{x} + l \int_\Omega \theta \, d\mathbf{x}, \tag{2.82}$$

and the minimum on the right hand side of (2.82) is attained for $\boldsymbol{\tau}_i = \mathbf{A}\nabla u_i$, $i = 1, \ldots, m$.

Then problem (2.9) reduces to

$$\min_{(\theta,\mathbf{A})\in\mathcal{A}} \min_{\substack{\boldsymbol{\tau}_i\in \mathrm{L}^2(\Omega;\mathbf{R}^d)\\ -\mathrm{div}\boldsymbol{\tau}_i=f_i}} \sum_{i=1}^m \mu_i \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau}_i\cdot\boldsymbol{\tau}_i\,d\mathbf{x} + l\int_\Omega \theta\,d\mathbf{x}. \tag{2.83}$$

Remark 2.10. Minimization problem (2.83) can be solved by the so-called *alternate direction method*, described in Allaire et al. (1997), which minimizes separately and successively in $\boldsymbol{\tau}_i$ and $(\theta,\mathbf{A})$.

In the case when the number of states is less then the space dimension, or in the spherically symmetric case, the alternate direction algorithm coincides with the optimality criteria method from Algorithm 2.17. Indeed, in k-th iteration of the alternate minimization algorithm for the problem (2.83), for fixed $(\theta^k,\mathbf{A}^k)$ the internal minimum is achieved for $\boldsymbol{\tau}_i = \boldsymbol{\sigma}_i^k = \mathbf{A}^k\nabla u_i^k$, where u_i^k is the state for $\mathbf{A}^k$, $i=1,\dots,m$. The update of the design variables is done by using optimality conditions (2.25) and (2.26) (where instead of $*$ we put k), and it coincides with the second step of Algorithm 2.17. We shall use this information in the proof of the convergence of our algorithm.

The convergence of Algorithm 2.7 (i.e. Algorithm 2.17) is stated below in Theorem 2.19. The proof goes along the same lines as the corresponding proof in the case of a single state equation (Allaire 2002, Theorem 5.1.5), which actually originates from the paper of Toader (Toader (1997)). However, we notice that, instead of using L^∞ weak-$*$ convergence of the (sub)sequence $(\mathbf{A}^k)$ of approximating designs, it is better and physically more relevant to use H-convergence. As the functional J is continuous with the respect to L^∞ weak-$*$ topology for θ and H-topology for $\mathbf{A}$, this abbreviates the proof. We shall use the following lemma (Allaire 2002, Lemma 3.1.9).

Lemma 2.18 The function $\phi:\mathbf{R}^+\times\mathbf{R}^d \longrightarrow \mathbf{R}$, defined by $\phi(a,\boldsymbol{\sigma}) = \frac{|\boldsymbol{\sigma}|^2}{a}$, is strictly convex and satisfies

$$\phi(a,\boldsymbol{\sigma}) = \phi(\bar a,\bar{\boldsymbol{\sigma}}) + \mathrm{D}\phi(\bar a,\bar{\boldsymbol{\sigma}})(a-\bar a,\boldsymbol{\sigma}-\bar{\boldsymbol{\sigma}}) + \phi(a,\boldsymbol{\sigma}-\frac{a}{\bar a}\bar{\boldsymbol{\sigma}}), \tag{2.84}$$

where the differential $D\phi$ is given by $\mathrm{D}\phi(\bar a,\bar{\boldsymbol{\sigma}})(b,\boldsymbol{\tau}) = -\frac{b}{\bar a^2}|\bar{\boldsymbol{\sigma}}|^2 + \frac{2}{\bar a}\bar{\boldsymbol{\sigma}}\cdot\boldsymbol{\tau}$.

□

Theorem 2.19 Let us consider the execution of Algorithm 2.7 in two particular cases: the spherically symmetric case with radial initial θ^0 and $\mathbf{A}^0\in\mathcal{K}(\theta^0)$ being a simple laminate with the lamination direction orthogonal to the radial vector, or the case when the number of states is less then the space dimension. Furthermore, let $(\theta^k,\mathbf{A}^k)$ be the sequence of iterations of Algorithm 2.7 for the minimization of the conic sum of energies, i.e. for the problem

$$\min_{(\theta,\mathbf{A})\in\mathcal{A}} J(\theta,\mathbf{A}) = \min_{(\theta,\mathbf{A})\in\mathcal{A}} \min_{\substack{\boldsymbol{\tau}_i\in \mathrm{L}^2(\Omega;\mathbf{R}^d)\\ -\mathrm{div}\boldsymbol{\tau}_i=f_i}} \sum_{i=1}^m \mu_i \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau}_i\cdot\boldsymbol{\tau}_i\,d\mathbf{x} + l\int_\Omega \theta\,d\mathbf{x}. \tag{2.85}$$

Then there exists a subsequence (denoted the same), which converges to a limit $(\widetilde{\theta},\widetilde{\mathbf{A}})$, in the sense that θ^k weak-$*$ converges in $\mathrm{L}^\infty(\Omega;[0,1])$ to $\widetilde{\theta}$, while $\mathbf{A}^k$ H-converges to $\widetilde{\mathbf{A}}$, and additionaly

$$\lim_{k\to\infty} J(\theta^k,\mathbf{A}^k) = J(\widetilde{\theta},\widetilde{\mathbf{A}}) = \min_{(\theta,\mathbf{A})\in\mathcal{A}} J(\theta,\mathbf{A}).$$

Proof. From the assumptions of the theorem, the iterations of Algorithm 2.7 coincide with those of Algorithm 2.17, and therefore with those of the alternate direction algorithm, by Remark 2.10. Moreover, without loss of generality we may assume that $l \le 0$, since otherwise, the optimality condition (2.26) implies that the optimum is achieved by using only a material with conductivity β. Since $(\theta^k,\mathbf{A}^k)$ belongs to $\mathcal{A}$, which is compact with respect to weak-$*$ topology for θ and H-topology for $\mathbf{A}$, we have (on a subsequence, that we denote the same)

$$\begin{aligned} \theta^k &\xrightarrow{*} \widetilde{\theta} \quad \text{in } \mathrm{L}^\infty(\Omega;[0,1]) \\ \mathbf{A}^k &\xrightarrow{H} \widetilde{\mathbf{A}} \quad \text{in } \mathrm{L}^\infty(\Omega;\mathcal{M}_{\alpha,\beta}). \end{aligned}$$

The definition of H-convergence implies

$$\boldsymbol{\sigma}_i^k = \mathbf{A}^k\nabla u_i^k \rightharpoonup \widetilde{\mathbf{A}}\nabla\widetilde{u}_i =: \widetilde{\boldsymbol{\sigma}}_i, \quad \text{in } \mathrm{L}^2(\Omega;\mathbf{R}^d), \quad i=1,\dots,m$$

where $\widetilde{u}_i$ is the state for $(\widetilde{\theta},\widetilde{\mathbf{A}})$. Moreover, from Lemma 2.2 the functional J is continuous on $\mathcal{A}$ with weak-$*$ topology for θ and H-topology for $\mathbf{A}$, and hence $J(\theta^k,\mathbf{A}^k) \longrightarrow J(\widetilde{\theta},\widetilde{\mathbf{A}})$. It remains to prove that $(\widetilde{\theta},\widetilde{\mathbf{A}})$ is a minimizer of the problem (2.85).

Let $(\theta^*,\mathbf{A}^*)$ be an optimal design for the functional J, and $\boldsymbol{\sigma}_i^* = \mathbf{A}^*\nabla u_i^*$, $i=1,\dots,m$, the corresponding minimizers of the inner minimization in (2.85). Using Lemma 2.18 with $\bar{\boldsymbol{\sigma}} = \sqrt{\mu_i}\boldsymbol{\sigma}_i^k$, $\bar{a} = \lambda^+_{\theta^{k+1}}$, $a = \lambda^+_{\theta^*}$, $\boldsymbol{\sigma} = \sqrt{\mu_i}\boldsymbol{\sigma}_i^*$, we get

$$\begin{aligned} J(\theta^*,\mathbf{A}^*) &= \int_\Omega \left(\sum_{i=1}^m \mu_i (\mathbf{A}^*)^{-1}\boldsymbol{\sigma}_i^*\cdot\boldsymbol{\sigma}_i^* + l\theta^*\right) d\mathbf{x} = \\ &= \int_\Omega \left(\frac{1}{\lambda^+_{\theta^*}}\sum_{i=1}^m \mu_i|\boldsymbol{\sigma}_i^*|^2 + l\theta^*\right) d\mathbf{x} = \\ &= \int_\Omega \left(\frac{1}{\lambda^+_{\theta^{k+1}}}\sum_{i=1}^m \mu_i|\boldsymbol{\sigma}_i^k|^2 + l\theta^{k+1}\right) d\mathbf{x} \\ &\quad + l\int_\Omega (\theta^* - \theta^{k+1})\,d\mathbf{x} + \sum_{i=1}^m \mu_i \int_\Omega \frac{\lambda^+_{\theta^{k+1}} - \lambda^+_{\theta^*}}{(\lambda^+_{\theta^{k+1}})^2}|\boldsymbol{\sigma}_i^k|^2 d\mathbf{x} \end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=1}^{m}\mu_i\int_\Omega \frac{2}{\lambda^+_{\theta^{k+1}}}\boldsymbol{\sigma}_i^k\cdot(\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k)\,d\mathbf{x}+\sum_{i=1}^{m}\mu_i\int_\Omega\frac{1}{\lambda^+_{\theta^*}}\left|\boldsymbol{\sigma}_i^*-\frac{\lambda^+_{\theta^*}}{\lambda^+_{\theta^{k+1}}}\boldsymbol{\sigma}_i^k\right|^2 d\mathbf{x}\\
\geq{}&\int_\Omega\left(\frac{1}{\lambda^+_{\theta^{k+1}}}\sum_{i=1}^{m}\mu_i|\boldsymbol{\sigma}_i^k|^2+l\theta^{k+1}\right)d\mathbf{x}+l\int_\Omega(\theta^*-\theta^{k+1})\,d\mathbf{x}\\
&+\sum_{i=1}^{m}\mu_i\int_\Omega\frac{(\beta-\alpha)(\theta^*-\theta^{k+1})}{(\lambda^+_{\theta^{k+1}})^2}|\boldsymbol{\sigma}_i^k|^2d\mathbf{x}+\sum_{i=1}^{m}\mu_i\int_\Omega\frac{2}{\lambda^+_{\theta^{k+1}}}\boldsymbol{\sigma}_i^k\cdot(\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k)\,d\mathbf{x}.
\end{aligned}
$$

From the optimality condition for θ^{k+1} (see Theorem 2.6), we have

$$
(\theta^*-\theta^{k+1})\left(\frac{\beta-\alpha}{(\lambda^+_{\theta^{k+1}})^2}\sum_{i=1}^{m}\mu_i|\boldsymbol{\sigma}_i^k|^2+l\right)\geq 0,\quad \text{a.e. } \mathbf{x}\in\Omega,
$$

which implies

$$
\begin{aligned}
J(\theta^*,\mathbf{A}^*)\geq{}&\int_\Omega\left(\frac{1}{\lambda^+_{\theta^{k+1}}}\sum_{i=1}^{m}\mu_i|\boldsymbol{\sigma}_i^k|^2+l\theta^{k+1}\right)d\mathbf{x}\\
&+\sum_{i=1}^{m}\mu_i\int_\Omega\frac{2}{\lambda^+_{\theta^{k+1}}}\boldsymbol{\sigma}_i^k\cdot(\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k)\,d\mathbf{x}.
\end{aligned}
\tag{2.86}
$$

On the other hand, if we define

$$
\boldsymbol{\sigma}_i^t:=(1-t)\boldsymbol{\sigma}_i^k+t\boldsymbol{\sigma}_i^*,\quad t\geq 0,
$$

we have

$$
\begin{aligned}
&\sum_{i=1}^{m}\mu_i\int_\Omega(\mathbf{A}^{k+1})^{-1}\boldsymbol{\sigma}_i^t\cdot\boldsymbol{\sigma}_i^t\,d\mathbf{x}=\\
&=\sum_{i=1}^{m}\mu_i\int_\Omega\Big((1-t)(\mathbf{A}^{k+1})^{-1}\boldsymbol{\sigma}_i^k+t(\mathbf{A}^{k+1})^{-1}\boldsymbol{\sigma}_i^*\Big)\cdot\Big((1-t)\boldsymbol{\sigma}_i^k+t\boldsymbol{\sigma}_i^*\Big)\,d\mathbf{x}\\
&=\sum_{i=1}^{m}\mu_i\Big(\int_\Omega(\mathbf{A}^{k+1})^{-1}\boldsymbol{\sigma}_i^k\cdot\boldsymbol{\sigma}_i^k\,d\mathbf{x}+2t\int_\Omega(\mathbf{A}^{k+1})^{-1}\boldsymbol{\sigma}_i^k\cdot(\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k)\,d\mathbf{x}\\
&\qquad+t^2\int_\Omega(\mathbf{A}^{k+1})^{-1}(\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k)\cdot(\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k)\,d\mathbf{x}\Big)\\
&\leq\sum_{i=1}^{m}\mu_i\int_\Omega\frac{|\boldsymbol{\sigma}_i^k|^2}{\lambda^+_{\theta^{k+1}}}\,d\mathbf{x}+t\sum_{i=1}^{m}\mu_i\int_\Omega\frac{2}{\lambda^+_{\theta^{k+1}}}\boldsymbol{\sigma}_i^k\cdot(\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k)\,d\mathbf{x}\\
&\qquad+\frac{t^2}{\alpha}\sum_{i=1}^{m}\mu_i\int_\Omega|\boldsymbol{\sigma}_i^*-\boldsymbol{\sigma}_i^k|^2\,d\mathbf{x}.
\end{aligned}
\tag{2.87}
$$

The above inequality arises from $\mathbf{A}^{k+1}\in \mathrm{L}^\infty(\Omega,\mathcal{M}_{\alpha,\beta})$, and $\mathbf{A}^{k+1}\boldsymbol{\sigma}_i^k=\lambda^+_{\theta^{k+1}}\boldsymbol{\sigma}_i^k$. Furthermore, since $l\leq 0$ and $\boldsymbol{\sigma}^{k+1}$ is optimal in the expression (2.82) for $J(\theta^{k+1},\mathbf{A}^{k+1})$, it follows

$$\begin{aligned} J(\theta^{k+1}, \mathbf{A}^{k+1}) &= \int_\Omega \left(\sum_{i=1}^m \mu_i (\mathbf{A}^{k+1})^{-1} \boldsymbol{\sigma}_i^{k+1} \cdot \boldsymbol{\sigma}_i^{k+1} + l\theta^{k+1} \right) d\mathbf{x} \\ &\leq \sum_{i=1}^m \mu_i \int_\Omega (\mathbf{A}^{k+1})^{-1} \boldsymbol{\sigma}_i^t \cdot \boldsymbol{\sigma}_i^t \, d\mathbf{x}. \end{aligned} \tag{2.88}$$

Using (2.86)–(2.88) after defining

$$\begin{aligned} I_k &:= \int_\Omega \left(\sum_{i=1}^m \mu_i (\mathbf{A}^{k+1})^{-1} \boldsymbol{\sigma}_i^k \cdot \boldsymbol{\sigma}_i^k + l\theta^{k+1} \right) d\mathbf{x} \\ &= \int_\Omega \left(\frac{1}{\lambda^+_{\theta^{k+1}}} \sum_{i=1}^m \mu_i |\boldsymbol{\sigma}_i^k|^2 + l\theta^{k+1} \right) d\mathbf{x}, \end{aligned}$$

we conclude

$$J(\theta^{k+1}, \mathbf{A}^{k+1}) - I_k \leq t(J(\theta^*, \mathbf{A}^*) - I_k) + \frac{t^2}{\alpha} \sum_{i=1}^m \mu_i \int_\Omega |\boldsymbol{\sigma}_i^* - \boldsymbol{\sigma}_i^k|^2 \, d\mathbf{x}.$$

Since the sequence $\boldsymbol{\sigma}_i^k$ is bounded in $\mathrm{L}^2(\Omega; \mathbf{R}^d)$, $i = 1, \dots, m$, there exists $c \geq 0$ such that

$$J(\theta^{k+1}, \mathbf{A}^{k+1}) - I_k \leq t(J(\theta^*, \mathbf{A}^*) - I_k) + ct^2, \quad t \geq 0, \ k \in \mathbf{N}.$$

After minimizing over $t \geq 0$ the right hand side of the above inequality, we get

$$\frac{(J(\theta^*, \mathbf{A}^*) - I_k)^2}{4c} \leq I_k - J(\theta^{k+1}, \mathbf{A}^{k+1}) \leq I_k - I_{k+1}, \tag{2.89}$$

where the last inequality follows from the property of the alternate minimization algorithm, which always decreases the value of the objective functional, i.e.

$$I_{k-1} \geq J(\theta^k, \mathbf{A}^k) \geq I_k.$$

Since the sequence I_k is decreasing and bounded below, it converges and clearly $I_k - I_{k-1} \longrightarrow 0$, as $k \longrightarrow \infty$. From (2.89) we easily conclude

$$J(\widetilde{\theta}, \widetilde{\mathbf{A}}) = \lim_{k \longrightarrow \infty} J(\theta^k, \mathbf{A}^k) = J(\theta^*, \mathbf{A}^*),$$

which proves the statement. □

Remark 2.11. Note that we have actually proved that any weakly convergent subsequence of $(\theta^k, \mathbf{A}^k)$ is a minimizing sequence for J, while existence of such subsequence of designs follows from the compactness of the corresponding topology.

Remark 2.12. As commented before, an exact solution for Example 2.3 is not known, but now we can confidently say that Algorithm 2.7 gives a converging sequence of designs for this problem as well.

Chapter 3

One state optimal design in linearized elasticity

The optimality criteria method developed in Section 2.2 for stationary diffusion equation will be adapted in this chapter to similar problems of optimal design in linearized elasticity. From Section 1.3, we recall the linearized elasticity system

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}e(\mathsf{u})) = \mathsf{f} & \text{in } \Omega \\ \mathsf{u} \in \mathrm{H}_0^1(\Omega;\mathbf{R}^d)\,, \end{cases} \tag{3.1}$$

where $\mathsf{f} \in \mathrm{H}^{-1}(\Omega;\mathbf{R}^{\mathrm{d}})$, $\mathbf{A} \in \mathrm{L}^\infty(\Omega;\mathcal{M}^4_{\alpha,\beta})$ and $e(\mathsf{u}) = \frac{1}{2}(\nabla\mathsf{u} + \nabla\mathsf{u}^\top)$. In problems of optimal design, we consider a specific stiffness tensor $\mathbf{A}$, made up of two isotropic elastic phases

$$\begin{aligned} \mathbf{A}_1 &= 2\mu_1\mathbf{I}_4 + \left(\kappa_1 - \frac{2\mu_1}{d}\right)\mathbf{I}_2\otimes\mathbf{I}_2 \\ \mathbf{A}_2 &= 2\mu_2\mathbf{I}_4 + \left(\kappa_2 - \frac{2\mu_2}{d}\right)\mathbf{I}_2\otimes\mathbf{I}_2, \end{aligned}$$

where $0 < \mu_1 \le \mu_2$ and $0 < \kappa_1 \le \kappa_2$. Let $\chi \in \mathrm{L}^\infty(\Omega;\{0,1\})$ denote a characteristic function of the part of the domain Ω occupied by $\mathbf{A}_1$. Then the overall stiffness tensor is given by

$$\mathbf{A}(\mathbf{x}) = \chi(\mathbf{x})\mathbf{A}_1 + (1-\chi(\mathbf{x}))\mathbf{A}_2, \quad \mathbf{x}\in\Omega.$$

A classical optimal design problem consists of minimizing the functional

$$J(\chi) = \int_\Omega \left[\chi(\mathbf{x})g_1(\mathbf{x},\mathsf{u}(\mathbf{x})) + (1-\chi(\mathbf{x}))g_2(\mathbf{x},\mathsf{u}(\mathbf{x}))\right]\,d\mathbf{x} \tag{3.2}$$

over the set of all measurable characteristic functions $\chi \in \mathrm{L}^\infty(\Omega;\{0,1\})$, with the prescribed amount $q \in \langle 0, |\Omega|\rangle$ of the material $\mathbf{A}_1$. Here, function u is the solution of the system (3.1), while functions g_1 and g_2 are some given functions which depend on the point $\mathbf{x}$ and on the value $\mathsf{u}(\mathbf{x})$ at the same point. If, for example, the objective functional is the compliance functional, by minimizing it we make the body as stiff as possible. Therefore, a minimum would be reached by using only the stronger phase $\mathbf{A}_2$. However, the constraint on the amount of the first material forces us to use both phases, which make this optimization problem highly nontrivial, and usually it does not admit a solution. In the relaxation process, we proceed similarly as in the conductivity case, using the homogenization method introduced in Section 1.3. If we impose that g_1, g_2 are Caratheodory functions satisfying the growth condition

$$|g_k(\mathbf{x}, \mathsf{u})| \leq a|\mathsf{u}|^s + b(\mathbf{x}), \quad k \in \{1, 2\},$$

for some $a > 0$, $b \in \mathrm{L}^1(\Omega)$ and $1 \leq s < \frac{2d}{d-2}$, then, the functional

$$J(\theta, \mathbf{A}) := \int_\Omega [\theta(\mathbf{x}) g_1(\mathbf{x}, \mathsf{u}(\mathbf{x})) + (1-\theta)(\mathbf{x})) g_2(\mathbf{x}, \mathsf{u}(\mathbf{x}))] \, d\mathbf{x} + l \int_\Omega \theta(\mathbf{x}) \, d\mathbf{x} \tag{3.3}$$

is continuous on a compact set $\mathrm{L}^\infty(\Omega; [0,1] \times \mathcal{M}^4_{\alpha,\beta})$ and the problem

$$\begin{cases} J(\theta, \mathbf{A}) \longrightarrow \min, \\ (\theta, \mathbf{A}) \in \mathcal{A} := \{(\theta, \mathbf{A}) \in \mathrm{L}^\infty(\Omega; [0,1] \times \mathcal{M}^4_{\alpha,\beta}) : \ \mathbf{A} \in \mathcal{K}(\theta) \text{ a.e. on } \Omega\} \end{cases} \tag{3.4}$$

is the relaxed formulation of the original problem (3.2) (see, for example, Allaire (2002)). The parameter l in (3.3) is a Lagrange multiplier introduced in order to handle the volume constraint. Remember that the set $\mathcal{A}$ is the set of all composite designs where, for each $\theta \in [0,1]$, $\mathcal{K}(\theta)$ is the set of all homogenized tensors obtained by mixing the phases $\mathbf{A}_1$ and $\mathbf{A}_2$ in proportions θ and $(1-\theta)$.

Theorem 3.1 The minimization problem (3.4) is a proper relaxation of the original problem (3.2) in the sense that

1. there exists at least one minimizer of J on $\mathcal{A}$,

2. up to a subsequence, every minimizing sequence of classical designs χ^n for J weak-$*$ converges in $\mathrm{L}^\infty(\Omega; [0,1])$ to a density function θ, and the associated stiffness tensor $\mathbf{A}^n = \chi^n \mathbf{A}_1 + (1-\chi^n)\mathbf{A}_2$ H-conveges to a composite stiffness tensor $\mathbf{A}$ such that $(\theta, \mathbf{A})$ is a minimizer of J on $\mathcal{A}$,

3. conversely, every minimizer $(\theta, \mathbf{A})$ of J on $\mathcal{A}$ can be attained as a limit of a minimizing sequence, for J, of classical designs χ^n, namely θ is the weak-$*$ limit of χ^n in $\mathrm{L}^\infty(\Omega; [0,1])$ and $\mathbf{A}$ is the H-limit of $\mathbf{A}^n = \chi^n \mathbf{A}_1 + (1-\chi^n)\mathbf{A}_2$.

Proof. The proof goes along the same lines as the proof of Theorem 2.3. □

Unlike in the conductivity setting, the G-closure set of all possible tensors obtained by mixing two isotropic elastic phases $\mathbf{A}_1$ and $\mathbf{A}_2$ in prescribed proportions is unknown. Lack of knowledge of the set $\mathcal{K}(\theta)$ prevents us from deriving the necessary conditions for

$(\theta^*, \mathbf{A}^*)$ to be optimal for (3.4). Therefore, it appears that the homogenization method is powerless. Nevertheless, there are some special cases of the objective functional where the set $\mathcal{K}(\theta)$ can be reduced to a smaller subset, which is explictly known, and therefore the homogenization theory can be used. One of those cases is the compliance optimization, which is studied in the next section.

3.1 Compliance minimization

In the sequel, we take $g_1 = g_2 = \mathsf{f} \cdot \mathsf{u}$ and the objective functional (3.3) reduces to

$$J(\theta, \mathbf{A}) = \int_\Omega \mathsf{f}(\mathbf{x}) \cdot \mathsf{u}(\mathbf{x})\, d\mathbf{x} + l \int_\Omega \theta(\mathbf{x})\, d\mathbf{x}. \tag{3.5}$$

The quantity $\int_\Omega \mathsf{f} \cdot \mathsf{u}\, d\mathbf{x}$ is called the compliance, the work done by the load. Therefore, by minimizing (3.5), one would like to find the most rigid structure, made of elastic materials $\mathbf{A}_1$ and $\mathbf{A}_2$.

Remark 3.1. The compliance function can itself be written as a minimization problem. To see this, first recall that a weak solution (1.46) of the linearized elasticity system (3.1) by the principle of complementary energy can equivalently be characterized as a unique minimizer of the quadratic functional

$$\frac{1}{2} \int_\Omega \mathbf{A} e(\mathsf{u}) : e(\mathsf{u})\, d\mathbf{x} - \int_\Omega \mathsf{f} \cdot \mathsf{u}\, d\mathbf{x}. \tag{3.6}$$

From the assumption $\mathbf{A}\boldsymbol{\xi} : \boldsymbol{\xi} \geq \alpha |\boldsymbol{\xi}|^2$, $\boldsymbol{\xi} \in \mathrm{Sym}_d$, for $\boldsymbol{\xi} = e(\mathsf{v}) - \mathbf{A}^{-1}\boldsymbol{\tau}$, $\mathsf{v} \in \mathrm{H}^1_0(\Omega; \mathbf{R}^d)$, $\boldsymbol{\tau} \in \mathrm{L}^2(\Omega; \mathrm{Sym}_d)$, monotonicity of the integral implies

$$\int_\Omega (\mathbf{A} e(\mathsf{v}) - \boldsymbol{\tau}) : (e(\mathsf{v}) - \mathbf{A}^{-1}\boldsymbol{\tau})\, d\mathbf{x} \geq 0, \quad \text{a.e. } \mathbf{x} \in \Omega, \tag{3.7}$$

which is equivalent to

$$-\int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau}\, d\mathbf{x} \leq \int_\Omega (\mathbf{A} e(\mathsf{v}) : e(\mathsf{v}) - 2\boldsymbol{\tau} : e(\mathsf{v}))\, d\mathbf{x}, \quad \text{a.e. } \mathbf{x} \in \Omega.$$

Equality in the above inequality is achieved if and only of $e(\mathsf{v}) - \mathbf{A}^{-1}\boldsymbol{\tau} = 0$ (since $\mathbf{A}$ is positive semidefinite), which happens if and only if $\boldsymbol{\tau} = \mathbf{A} e(\mathsf{v})$. By maximizing in the last inequality over the set of all $\boldsymbol{\tau} \in \mathrm{L}^2(\Omega; \mathrm{Sym}_d)$ such that $-\mathsf{div}\, \boldsymbol{\tau} = \mathsf{f}$ in Ω, we have

$$-\min_{\substack{\boldsymbol{\tau} \in \mathrm{L}^2(\Omega; \mathrm{Sym_d}) \\ -\mathrm{div}\boldsymbol{\tau} = \mathsf{f} \text{ in } \Omega}} \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau} = \int_\Omega (\mathbf{A} e(\mathsf{v}) : e(\mathsf{v}) - 2\mathsf{f} \cdot \mathsf{v})\, d\mathbf{x}, \quad \mathsf{v} \in \mathrm{H}^1_0(\Omega; \mathbf{R}^d).$$

In particular, for the solution u of (3.1), it follows

$$-\int_\Omega \mathsf{f} \cdot \mathsf{u}\, d\mathbf{x} = \int_\Omega (\mathbf{A} e(\mathsf{u}) : e(\mathsf{u}) - 2\mathsf{f} \cdot \mathsf{u})\, d\mathbf{x} = -\min_{\substack{\boldsymbol{\tau} \in \mathrm{L}^2(\Omega; \mathrm{Sym_d}) \\ -\mathrm{div}\boldsymbol{\tau} = \mathsf{f} \text{ in } \Omega}} \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau}\, d\mathbf{x}.$$

Therefore, the functional (3.5) becomes

$$J(\theta, \mathbf{A}) = \min_{\substack{\boldsymbol{\tau}\in \mathrm{L}^2(\Omega;\mathrm{Sym_d}) \\ -\mathrm{div}\boldsymbol{\tau}=\mathsf{f} \text{ in } \Omega}} \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau}\, d\mathbf{x} + l\int_\Omega \theta\, d\mathbf{x}, \tag{3.8}$$

where the minimum on the left side is achieved by $\boldsymbol{\tau} = \mathbf{A}e(\mathsf{u})$.

According to the above remark, the relaxed problem (3.4) can be considered as a double minimization in $(\theta, \mathbf{A})$ and in $\boldsymbol{\tau}$. Since the sets by which the minimization is performed are independent, we obtain

$$\min_{(\theta,\mathbf{A})\in\mathcal{A}} J(\theta, \mathbf{A}) = \min_{\substack{\boldsymbol{\tau}\in \mathrm{L}^2(\Omega;\mathrm{Sym_d}) \\ -\mathrm{div}\boldsymbol{\tau}=\mathsf{f} \text{ in } \Omega}} \min_{(\theta,\mathbf{A})\in\mathcal{A}} \int_\Omega (\mathbf{A}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau} + l\theta)\, d\mathbf{x}. \tag{3.9}$$

This approach is suitable for deriving the neccessary conditions of optimality, stated in the following theorem.

Theorem 3.2 If $(\theta^*, \mathbf{A}^*)$ is a minimizer of the objective function (3.5), and if $\boldsymbol{\sigma}^*$ is the unique corresponding minimizer in (3.8), then $\boldsymbol{\sigma}^* = \mathbf{A}^*e(\mathsf{u}^*)$, where u^* is the state function for $(\theta^*, \mathbf{A}^*)$. Furthermore, $\mathbf{A}^*$ satisfies, almost everywhere in Ω,

$$\mathbf{A}^{*-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* = h(\theta^*, \boldsymbol{\sigma}^*), \tag{3.10}$$

where $h(\theta^*, \boldsymbol{\sigma}^*)$ is the lower Hashin-Shtrikman bound on the complementary energy defined by (1.61), i.e.

$$h(\theta, \boldsymbol{\sigma}) = \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta \max_{\boldsymbol{\eta}\in\mathrm{Sym_d}} \left[2\boldsymbol{\sigma} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta)g^c(\boldsymbol{\eta})\right], \tag{3.11}$$

where $g^c(\boldsymbol{\eta}) = \max_{\mathsf{e}\in S^{d-1}} (f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta})$ and

$$f_2^c(\mathsf{e})\boldsymbol{\xi} : \boldsymbol{\xi} = \mathbf{A}_2\boldsymbol{\xi} : \boldsymbol{\xi} - \frac{1}{\mu_2}|\mathbf{A}_2\boldsymbol{\xi}\mathsf{e}|^2 + \frac{\mu_2+\lambda_2}{\mu_2(2\mu_2+\lambda_2)}\left((\mathbf{A}_2\boldsymbol{\xi})\mathsf{e}\cdot\mathsf{e}\right)^2, \quad \boldsymbol{\xi}\in\mathrm{Sym_d}, \tag{3.12}$$

and θ^* is the unique minimizer of the convex minimization problem

$$\min_{0\le\theta\le1} (h(\theta, \boldsymbol{\sigma}^*) + l\theta), \quad \text{a.e. on } \Omega. \tag{3.13}$$

Proof. Let $(\theta^*, \mathbf{A}^*)$ be a minimizer of the objective function (3.5), and $\boldsymbol{\sigma}^*$ the corresponding minimizer in (3.8). From Remark 3.1 it is clear that $\boldsymbol{\sigma}^* = \mathbf{A}^*e(\mathsf{u}^*)$, where u^* is the state function for $(\theta^*, \mathbf{A}^*)$. In view of (3.9), $(\theta^*, \mathbf{A}^*)$ is also a minimizer of

$$\min_{(\theta,\mathbf{A})\in\mathcal{A}} \int_\Omega \left(\mathbf{A}^{-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* + l\theta\right)\, d\mathbf{x}.$$

By the local character of G-closure, i.e. since

$$\mathcal{A} = \{(\theta, \mathbf{A}) \in \mathrm{L}^\infty(\Omega; [0,1] \times \mathcal{M}^4_{\alpha,\beta}) : \mathbf{A}(\mathbf{x}) \in \mathcal{K}(\theta(\mathbf{x})) \,\text{a.e. } \mathbf{x} \in \Omega\},$$

the following is valid:

$$\begin{aligned} &\min_{(\theta,\mathbf{A})\in\mathcal{A}} \int_\Omega \left(\mathbf{A}^{-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta(\mathbf{x})\right) d\mathbf{x} = \\ &= \int_\Omega \min_{0\le\theta\le1} \left(\min_{\mathbf{A}\in G_\theta} \mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta\right) d\mathbf{x}. \end{aligned} \tag{3.14}$$

Indeed, let $(\tilde{\theta}, \tilde{\mathbf{A}}) \in \mathcal{A}$. For fixed $\mathbf{x} \in \Omega$ we have

$$\tilde{\mathbf{A}}^{-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\tilde{\theta}(\mathbf{x}) \ge \min_{0\le\theta\le1} \left(\min_{\mathbf{A}\in G_\theta} \mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta\right)$$

and integration over Ω leads to

$$\int_\Omega \left(\tilde{\mathbf{A}}^{-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\tilde{\theta}(\mathbf{x})\right) d\mathbf{x} \ge \int_\Omega \min_{0\le\theta\le1} \left(\min_{\mathbf{A}\in G_\theta} \mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta\right) d\mathbf{x}.$$

In particular, the above inequality is valid for $(\theta^*, \mathbf{A}^*)$, and it follows that

$$\min_{(\theta,\mathbf{A})\in\mathcal{A}} \int_\Omega \left(\mathbf{A}^{-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta(\mathbf{x})\right) d\mathbf{x} \ge \int_\Omega \min_{0\le\theta\le1} \left(\min_{\mathbf{A}\in G_\theta} \mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta\right) d\mathbf{x}.$$

For the reverse inequality, we fix $\mathbf{x} \in \Omega$, and define $\tilde{\theta}(\mathbf{x})$ as $\theta \in [0,1]$ for which the external minimum on the right-hand side in (3.14) is achieved. Moreover, define $\tilde{\mathbf{A}}(\mathbf{x})$ as $\mathbf{A} \in G_{\tilde{\theta}(\mathbf{x})}$ for which the internal minimum on the right-hand side of (3.14) is achieved. Then

$$\tilde{\mathbf{A}}^{-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\tilde{\theta}(\mathbf{x}) = \min_{0\le\theta\le1} \left(\min_{\mathbf{A}\in G_\theta} \mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta\right),$$

and integration over Ω gives

$$\begin{aligned} \min_{(\theta,\mathbf{A})\in\mathcal{A}} \int_\Omega \left(\mathbf{A}^{-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta(\mathbf{x})\right) d\mathbf{x} &\le \int_\Omega \left(\tilde{\mathbf{A}}^{-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\tilde{\theta}(\mathbf{x})\right) d\mathbf{x} = \\ &= \int_\Omega \min_{0\le\theta\le1} \left(\min_{\mathbf{A}\in G_\theta} \mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta\right) d\mathbf{x}, \end{aligned}$$

which proves (3.14). Furthermore, equality (3.14) implies that $(\theta^*(\mathbf{x}), \mathbf{A}^*(\mathbf{x}))$ is a minimizer for

$$\min_{0\le\theta\le1} \left(\min_{\mathbf{A}\in G_\theta} \mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) + l\theta\right), \quad \text{a.e. } \mathbf{x} \in \Omega.$$

Therefore, from Proposition 1.21 it follows

$$\mathbf{A}^{*-1}(\mathbf{x})\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x}) = \min_{\mathbf{A}\in G_{\theta^*(\mathbf{x})}} \left(\mathbf{A}^{-1}\boldsymbol{\sigma}^*(\mathbf{x}) : \boldsymbol{\sigma}^*(\mathbf{x})\right) = h(\theta^*(\mathbf{x}), \boldsymbol{\sigma}^*(\mathbf{x})) \quad \text{a.e. } \mathbf{x} \in \Omega,$$

where $h(\theta^*, \boldsymbol{\sigma}^*)$ is the lower Hashin-Shtrikman bound on complementary energy (3.11). This gives the condition (3.10). Moreover, function $h(\theta, \boldsymbol{\sigma})$, for $\theta \in [0,1]$ and $\boldsymbol{\sigma} \in \mathrm{Sym}_d$, is a C^1 function with the respect to θ, and a strictly convex function of θ, since it is quadratic function in θ and $g^c(\boldsymbol{\eta})$ is strictly positive for $\boldsymbol{\sigma} \neq \mathbf{0}$. Hence, the function

$$\theta \mapsto h(\theta, \boldsymbol{\sigma}^*(\mathbf{x})) + l\theta$$

admits a unique minimizer $\theta^*(\mathbf{x})$ in $[0,1]$, for almost every $\mathbf{x} \in \Omega$ and the proof is complete. $\square$

Remark 3.2. The optimality conditions derived in Theorem 3.2 are valid if we consider a problem with mixed boundary conditions, i.e. the problem

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}e(\mathsf{u})) = \mathsf{f} & \text{in } \Omega \\ \mathsf{u} = \mathbf{0} & \text{on } \Gamma_D \\ \mathbf{A}e(\mathsf{u})\mathsf{n} = \mathsf{g} & \text{on } \Gamma_N, \end{cases} \tag{3.15}$$

instead of (3.1), with Γ_D nonempty and $\Gamma_D \cup \Gamma_N = \partial\Omega$. In this case we can repeat steps from Remark 3.1, starting from the fact that the weak solution of (3.15) can equivalently be characterized as the unique minimizer of the functional

$$\frac{1}{2}\int_\Omega \mathbf{A}e(\mathsf{u}) : e(\mathsf{u})\, d\mathbf{x} - \int_\Omega \mathsf{f}\cdot\mathsf{u}\, d\mathbf{x} - \int_{\Gamma_N} \mathsf{g}\cdot\mathsf{u}\, dS.$$

The compliance functional in this case reads

$$\int_\Omega \mathsf{f}\cdot\mathsf{u}\, d\mathbf{x} + \int_{\Gamma_N} \mathsf{g}\cdot\mathsf{u}\, dS = \min_{\substack{\boldsymbol{\tau}\in \mathrm{L}^2(\Omega;\mathrm{Sym_d}) \\ -\mathrm{div}\boldsymbol{\tau}=\mathsf{f} \text{ in } \Omega \\ \boldsymbol{\tau}\mathsf{n}=\mathsf{g} \text{ on } \Gamma_N}} \int_\Omega \mathbf{A}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau}\, d\mathbf{x}. \tag{3.16}$$

From Theorem 3.2 we can deduce that in the case of compliance minimization, one does not need to know the entire $\mathcal{K}(\theta)$ in order to find the optimal pair $(\theta^*, \mathbf{A}^*)$, but only its subset of sequential laminates which saturate the lower Hashin-Shtrikman bound (3.11). More precisely, if we define the set $\mathcal{L}^+$ of sequentially laminated designs as

$$\mathcal{L}^+ := \left\{(\theta, \mathbf{A}) \in \mathrm{L}^\infty(\Omega; [0,1] \times \mathcal{M}^4_{\alpha,\beta}) : \mathbf{A}(\mathbf{x}) \in L^+_{\theta(\mathbf{x})},\ \text{a.e. in } \Omega\right\},$$

where the set L^+_θ, for $\theta \in [0,1]$, is the set of all sequential laminates $\mathbf{A}$, with core $\mathbf{A}_1$ and matrix $\mathbf{A}_2$, in proportions θ and $(1-\theta)$, respectively, introduced in Definition 1.5, then the following theorem holds.

Theorem 3.3 For the objective function (3.5) we have

$$\min_{(\theta,\mathbf{A})\in\mathcal{A}} J(\theta, \mathbf{A}) = \min_{(\theta,\mathbf{A})\in\mathcal{L}^+} J(\theta, \mathbf{A}).$$

If $(\theta^*, \mathbf{A}^*)$ is a minimizer of J in $\mathcal{A}$, and if $\boldsymbol{\sigma}^*$ is its associated stress tensor which minimizes (3.8), then there exists a sequential laminate $\tilde{\mathbf{A}}$ such that $(\theta^*, \tilde{\mathbf{A}})$ is a minimizer of J in $\mathcal{L}^+$, $\boldsymbol{\sigma}^*$ is again its associated stress tensor, and $\mathbf{A}^{*-1}\boldsymbol{\sigma}^* = \tilde{\mathbf{A}}^{-1}\boldsymbol{\sigma}^*$. Furthermore, $\tilde{\mathbf{A}}$ can be chosen among rank-d sequential laminates (with core $\mathbf{A}_1$ and matrix $\mathbf{A}_2$) having the same lamination directions as the orthogonal basis of the eigenvectors of $\boldsymbol{\sigma}^*$.

Proof. Let $(\theta^*, \mathbf{A}^*)$ be a minimizer of J on $\mathcal{A}$, and $\boldsymbol{\sigma}^*$ be its associated stress tensor which minimizes (3.8). By Theorem 3.2

$$\mathbf{A}^{*-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* = h(\theta^*, \boldsymbol{\sigma}^*) = \min_{\mathbf{A}\in G_{\theta^*}} \mathbf{A}^{-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* \quad \text{a.e. on } \Omega. \tag{3.17}$$

Moreover, by Proposition 1.26, there exists a rank-d sequential laminate $\tilde{\mathbf{A}}$ with lamination direction given by eigenvectors of $\boldsymbol{\sigma}^*$ which saturates the lower Hashin-Shtrikman bound (1.61), i.e.

$$\tilde{\mathbf{A}}^{-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* = \min_{\mathbf{A}\in G_{\theta^*}} \mathbf{A}^{-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* \quad \text{a.e. on } \Omega. \tag{3.18}$$

Using (3.17) and (3.18), it follows

$$J(\theta^*, \mathbf{A}^*) = \int_\Omega \mathbf{A}^{*-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* \, d\mathbf{x} + l \int_\Omega \theta^* \, d\mathbf{x} = \int_\Omega \tilde{\mathbf{A}}^{-1}\boldsymbol{\sigma}^* : \boldsymbol{\sigma}^* \, d\mathbf{x} + l \int_\Omega \theta^* \, d\mathbf{x} = J(\theta^*, \tilde{\mathbf{A}}).$$

From the above equality we conclude that $(\theta^*, \tilde{\mathbf{A}}) \in \mathcal{L}^+$ is also a minimizer for J on $\mathcal{A}$ and using (3.8) we conclude that $\boldsymbol{\sigma}^*$ is the stress tensor associated to $(\theta^*, \tilde{\mathbf{A}})$, which minimizes

$$\min_{\substack{\boldsymbol{\tau}\in \mathrm{L}^2(\Omega;\mathrm{Sym}_d) \\ -\mathrm{div}\boldsymbol{\tau}=\mathsf{f} \text{ in } \Omega}} \int_\Omega \tilde{\mathbf{A}}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau} \, d\mathbf{x}.$$

It remains to prove $\mathbf{A}^{*-1}\boldsymbol{\sigma}^* = \tilde{\mathbf{A}}^{-1}\boldsymbol{\sigma}^*$. Consider the function

$$\boldsymbol{\tau} \mapsto \mathbf{A}^{*-1}\boldsymbol{\tau} : \boldsymbol{\tau} - h(\theta^*, \boldsymbol{\tau}). \tag{3.19}$$

Recall that the function $h(\theta^*, \boldsymbol{\tau})$ is given as the extremal value of a strictly concave function of $\boldsymbol{\eta}$, which is extremal at a unique $\boldsymbol{\eta}^*$. Thus, when differentiating it, the derivative of $\boldsymbol{\eta}^*$ cancels out due to its optimality condition, which implies that $h(\theta^*, \boldsymbol{\tau})$ is C^1 function with the respect to $\boldsymbol{\tau}$. Furthermore, (3.19) is a nonnegative function which vanishes for $\boldsymbol{\tau} = \boldsymbol{\sigma}^*$, implying that it has a minimum at $\boldsymbol{\tau} = \boldsymbol{\sigma}^*$, and the necessary condition of optimality implies

$$\mathbf{A}^{*-1}\boldsymbol{\sigma}^* = \frac{\partial h}{\partial \boldsymbol{\tau}}(\theta^*, \boldsymbol{\sigma}^*).$$

Since $\tilde{\mathbf{A}}$ is also optimal, we can repeat this procedure for the function $\boldsymbol{\tau} \mapsto \tilde{\mathbf{A}}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau} - h(\theta^*, \boldsymbol{\tau})$, which then gives $\mathbf{A}^{*-1}\boldsymbol{\sigma}^* = \tilde{\mathbf{A}}^{-1}\boldsymbol{\sigma}^*$. □

Theorem 3.3 justifies the use of the homogenization method for compliance minimization. Instead of minimizing compliance on the set of all composite materials, which is unknown, the minimization can be performed on its subset of sequential laminates, which is fully described (see Section 1.3).

3.1.1 Optimality criteria method

Using the neccessary conditions of optimality from the Theorem 3.2, an optimality criteria method can be derived. The algorithm is stated below.

Algorithm 3.4 Take some initial θ^0 and $\mathbf{A}^0$. For $k \geq 0$:

(1) Calculate u^k, the solution of

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}^k e(\mathsf{u}^k)) = \mathsf{f} & \text{in } \Omega \\ \mathsf{u}^k \in \mathrm{H}_0^1(\Omega; \mathbf{R}^d)\,, \end{cases}$$

and define $\boldsymbol{\sigma}^k := \mathbf{A}^k e(\mathsf{u}^k)$.
If $\boldsymbol{\sigma}^k(\mathbf{x}) = \mathbf{0}$, for some $\mathbf{x} \in \Omega$, then leave the old data for the next iteration of $\theta^{k+1}(\mathbf{x})$ and $\mathbf{A}^{k+1}(\mathbf{x})$. Else do step 2.

(2) For $\mathbf{x} \in \Omega$, take $\theta^{k+1}(\mathbf{x})$ as a zero of the function

$$\theta \mapsto \frac{\partial h}{\partial \theta}(\theta, \boldsymbol{\sigma}^k(\mathbf{x})) + l\,,$$

and if a zero doesn't exist, take 0 (or 1) if the function is positive (or negative) on $[0,1]$.
Find the eigenvalues and eigenvectors of the matrix $\boldsymbol{\sigma}^k(\mathbf{x})$ and let $\mathbf{A}^{k+1}(\mathbf{x})$ be a sequential laminate which saturates the lower Hashin-Shtrikman bound (3.11). In particular, $\mathbf{A}^{k+1}(\mathbf{x})$ is taken from the set $L^+_{\theta^{k+1}(\mathbf{x})}$ with lamination directions that are from the orthogonal basis of the eigenvectors of $\boldsymbol{\sigma}^k(\mathbf{x})$.

As in the case of energy minimization in the conductivity case, Algorithm 3.4 coincides with the alternate direction algorithm, which amounts to minimizing (3.9) iteratively and separately in $\boldsymbol{\tau}$ and $(\theta, \mathbf{A})$.

In order to implement the algorithm, one must know the lower Hashin-Shtrikman bound on complementary energy explictly. The extremal e in the definition of the function $g^c(\boldsymbol{\eta})$ in (3.11) is the eigenvector of $\boldsymbol{\eta}$ associated to its smallest eigenvalue by the absolute value (see proof of Lemma 1.22), and explicit formula for the $g^c(\boldsymbol{\eta})$ is given in Lemma 1.22. Explicit computation of the function $h(\theta, \boldsymbol{\sigma})$ is given in Gibianski & Cherkaev (1984) and Allaire & Kohn (1993*b*) for two and in Gibianski & Cherkaev (1987) and Allaire (1994) for three dimensional case, but only for shape optimization, where one phase is replaced by holes.

In the following theorem we give the explicit lower Hashin-Shtrikman bound on complementary energy in two space dimensions and the corresponding optimal microstructure.

Theorem 3.5 If dimension $d = 2$, then for given $\theta \in [0,1]$, $0 < \mu_1 < \mu_2$, $0 < \kappa_1 < \kappa_2$ and matrix $\boldsymbol{\sigma}$ with eigenvalues σ_1 and σ_2, the lower Hashin-Shtrikman bound on complementary energy

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta \max_{\boldsymbol{\eta} \in \mathrm{Sym}_2} \left[2\boldsymbol{\sigma} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta) g^c(\boldsymbol{\eta}) \right], \quad (3.20)$$

is explicitly given as follows:

A. If

$$\begin{aligned} &(1-\theta)\mu_2(\kappa_2-\kappa_1)|\sigma_1+\sigma_2| < \Big(\kappa_1(\kappa_2+\mu_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)|\sigma_1-\sigma_2|, \\ &(1-\theta)\kappa_2(\mu_2-\mu_1)|\sigma_1-\sigma_2| < \Big(\mu_1(\kappa_2+\mu_2) + (1-\theta)\kappa_2(\mu_2-\mu_1)\Big)|\sigma_1+\sigma_2|, \end{aligned} \quad (3.21)$$

then

$$\begin{aligned} \mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} &\geq \theta \mathbf{A}_1^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + (1-\theta)\mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} - \\ &- (1-\theta)\theta \frac{(\kappa_2+\mu_2)\Big(\kappa_1\kappa_2(\mu_2-\mu_1)|\sigma_1-\sigma_2| + \mu_1\mu_2(\kappa_2-\kappa_1)|\sigma_1+\sigma_2|\Big)^2}{4\kappa_2\mu_2\Big(\kappa_1\mu_1(\mu_2+\kappa_2) + (1-\theta)\Big(\mu_1\mu_2(\kappa_2-\kappa_1) + \kappa_1\kappa_2(\mu_2-\mu_1)\Big)\Big)}. \end{aligned}$$

This bound is achieved by a simple laminate with the lamination direction orthogonal to an eigenvector associated to the eigenvalue of the smallest absolute value of extremal $\boldsymbol{\eta}$ in (3.20).

B. If

$$(1-\theta)\mu_2(\kappa_2-\kappa_1)|\sigma_1+\sigma_2| \geq \Big(\kappa_1(\kappa_2+\mu_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)|\sigma_1-\sigma_2|, \quad (3.22)$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{\theta(\kappa_2-\kappa_1)(\kappa_2+\mu_2)(\sigma_1+\sigma_2)^2}{4\kappa_2\Big((1-\theta)\kappa_2\mu_2 + \kappa_1(\kappa_2+\theta\mu_2)\Big)}. \quad (3.23)$$

This bound can be achieved by the rank-2 sequential laminate with the lamination directions given by the eigenvectors e_1 and e_2 of extremal $\boldsymbol{\eta}$ in (3.20), and lamination parameters

$$\begin{aligned} m_1 &= \frac{1}{2} + \frac{\Big(\kappa_1(\mu_2+\kappa_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)(\sigma_2-\sigma_1)}{2(1-\theta)(\kappa_2-\kappa_1)\mu_2(\sigma_1+\sigma_2)} \\ m_2 &= \frac{1}{2} + \frac{\Big(\kappa_1(\mu_2+\kappa_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)(\sigma_1-\sigma_2)}{2(1-\theta)(\kappa_2-\kappa_1)\mu_2(\sigma_1+\sigma_2)}. \end{aligned} \quad (3.24)$$

C. If

$$(1-\theta)\kappa_2(\mu_2-\mu_1)|\sigma_1-\sigma_2| \geq \Big(\mu_1(\kappa_2+\mu_2)+(1-\theta)\kappa_2(\mu_2-\mu_1)\Big)|\sigma_1+\sigma_2|, \quad (3.25)$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma} + \frac{\theta(\mu_2-\mu_1)(\kappa_2+\mu_2)(\sigma_1-\sigma_2)^2}{4\mu_2\Big((1-\theta)\kappa_2\mu_2+\mu_1(\mu_2+\theta\kappa_2)\Big)}. \quad (3.26)$$

In this case, the bound can be achieved by the rank-2 sequential laminate with the lamination direction given by the eigenvectors e_1 and e_2 of the extremal $\boldsymbol{\eta}$ in (3.20), and lamination parameters

$$\begin{aligned} m_1 &= \frac{1}{2} + \frac{\Big(\mu_1(\mu_2+\kappa_2)+(1-\theta)\kappa_2(\mu_2-\mu_1)\Big)(\sigma_1+\sigma_2)}{2(1-\theta)(\mu_2-\mu_1)\kappa_2(\sigma_2-\sigma_1)} \\ m_2 &= \frac{1}{2} + \frac{\Big(\mu_1(\mu_2+\kappa_2)+(1-\theta)\kappa_2(\mu_2-\mu_1)\Big)(\sigma_1+\sigma_2)}{2(1-\theta)(\mu_2-\mu_1)\kappa_2(\sigma_1-\sigma_2)}. \end{aligned} \quad (3.27)$$

Proof. We prove this theorem in two parts. First we explicitly calculate the lower Hashin-Shtrikman bound and then we find the optimal microstructure. The proof is similar to the one from Allaire & Kohn (1993*a*) for the explicit primal Hashin-Shtrikman bounds.

In order to find the explicit bound, we need to solve the maximization problem

$$\max_{\boldsymbol{\eta}\in\mathrm{Sym}_2}\Big[2\boldsymbol{\sigma}:\boldsymbol{\eta} - (\mathbf{A}_1^{-1}-\mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta}:\boldsymbol{\eta} - (1-\theta)g^c(\boldsymbol{\eta})\Big], \quad (3.28)$$

where $g^c(\boldsymbol{\eta})$ is given in Lemma 1.22, for $d=2$. Let us introduce a notation $\delta\mu=\mu_2-\mu_1$ and $\delta\kappa=\kappa_2-\kappa_1$. By Theorem 2.5, the maximum of $\boldsymbol{\sigma}:\boldsymbol{\eta}$ is obtained when matrices $\boldsymbol{\sigma}$ and $\boldsymbol{\eta}$ are simultaneously diagonalizable. Moreover, since $\mathbf{A}_1$ and $\mathbf{A}_2$ are isotropic, it follows

$$(\mathbf{A}_1^{-1}-\mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta}:\boldsymbol{\eta} = \frac{2\mu_1\mu_2}{\delta\mu}(\eta_1^2+\eta_2^2) + \left(\frac{\kappa_1\kappa_2}{\delta\kappa} - \frac{2\mu_1\mu_2}{3\delta\mu}\right)(\eta_1+\eta_2)^2,$$

where η_1 and η_2 are eigenvalues of $\boldsymbol{\eta}$. Therefore, the maximization over all symmetric 2×2 matrices in (3.28) is equivalent to the two dimensional maximization of the concave function

$$F(\eta_1,\eta_2) = 2(\sigma_1\eta_1+\sigma_2\eta_2) - \frac{2\mu_1\mu_2}{\delta\mu}(\eta_1^2+\eta_2^2) - \left(\frac{\kappa_1\kappa_2}{\delta\kappa} - \frac{2\mu_1\mu_2}{3\delta\mu}\right)(\eta_1+\eta_2)^2 - (1-\theta)g^c(\eta_1,\eta_2), \quad (3.29)$$

over all pairs $(\eta_1,\eta_2)\in\mathbf{R}^2$, where

$$\begin{aligned} g^c(\eta_1,\eta_2) &= 2\mu_2(\eta_1^2+\eta_2^2)+\lambda_2(\eta_1+\eta_2)^2 - \\ &\quad - \frac{1}{2\mu_2+\lambda_2}\min\left\{(2\mu_2\eta_1+\lambda_2(\eta_1+\eta_2))^2, (2\mu_2\eta_2+\lambda_2(\eta_1+\eta_2))^2\right\}. \end{aligned}$$

Note that the function $g^c(\eta_1,\eta_2)$ is continuously differentiable except when $(2\mu_2\eta_1+\lambda_2(\eta_1+\eta_2))^2 = (2\mu_2\eta_2+\lambda_2(\eta_1+\eta_2))^2$, i.e. on the lines $\eta_1=\eta_2$ and $\eta_1=-\eta_2$. Therefore, we shall consider three cases: $|\eta_1|<|\eta_2|$, $|\eta_1|>|\eta_2|$ and $|\eta_1|=|\eta_2|$.

I. Let us assume that $|\eta_1|<|\eta_2|$. This implies that

$$g^c(\eta_1,\eta_2) = 2\mu_2(\eta_1^2+\eta_2^2)+\lambda_2(\eta_1+\eta_2)^2 - \frac{1}{2\mu_2+\lambda_2}(2\mu_2\eta_1+\lambda_2(\eta_1+\eta_2))^2,$$

and from the necessary and sufficient condition of optimality, $\nabla F = 0$, we get

$$\begin{aligned}
\sigma_1\delta\kappa\delta\mu &= (\delta\kappa\mu_1\mu_2+\delta\mu\kappa_1\kappa_2)\eta_1 + (\delta\mu\kappa_1\kappa_2-\delta\mu_1\mu_2)\eta_2\\
\sigma_2\delta\kappa\delta\mu(\mu_2+\kappa_2) &= (\mu_2+\kappa_2)(\delta\mu\kappa_1\kappa_2-\delta\mu_1\mu_2)\eta_1+\\
&\quad+\Big((\mu_2+\kappa_2)(\delta\kappa\mu_1\mu_2+\delta\mu\kappa_1\kappa_2)+4(1-\theta)\delta\kappa\delta\mu\kappa_2\mu_2\Big)\eta_2.
\end{aligned} \tag{3.30}$$

The solution of the above system is unique:

$$\begin{aligned}
\eta_1 &= \frac{\delta\kappa\mu_2\Big(\mu_1(\mu_2+\kappa_2)+2(1-\theta)\delta\mu\kappa_2\Big)(\sigma_1+\sigma_2)}{4\mu_2\kappa_2\Big(\kappa_1\mu_1(\mu_2+\kappa_2)+(1-\theta)(\delta\kappa\mu_1\mu_2+\delta\mu\kappa_1\kappa_2)\Big)}+\\
&\quad+\frac{\delta\mu\kappa_2\Big(\kappa_1(\mu_2+\kappa_2)+2(1-\theta)\delta\kappa\mu_2\Big)(\sigma_1-\sigma_2)}{4\mu_2\kappa_2\Big(\kappa_1\mu_1(\mu_2+\kappa_2)+(1-\theta)(\delta\kappa\mu_1\mu_2+\delta\mu\kappa_1\kappa_2)\Big)}\\
\eta_2 &= \frac{(\mu_2+\kappa_2)\Big(\delta\kappa\mu_1\mu_2(\sigma_1+\sigma_2)-\delta\mu\kappa_1\kappa_2(\sigma_1-\sigma_2)\Big)}{4\mu_2\kappa_2\Big(\kappa_1\mu_1(\mu_2+\kappa_2)+(1-\theta)(\delta\kappa\mu_1\mu_2+\delta\mu\kappa_1\kappa_2)\Big)},
\end{aligned} \tag{3.31}$$

and therefore, (3.20) becomes

$$\begin{aligned}
&\mathbf{A}^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma} \geq \theta\mathbf{A}_1^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}+(1-\theta)\mathbf{A}_2^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}-\\
&\quad-(1-\theta)\theta\frac{(\kappa_2+\mu_2)\Big(\kappa_1\kappa_2(\mu_2-\mu_1)(\sigma_1-\sigma_2)+\mu_1\mu_2(\kappa_2-\kappa_1)(\sigma_1+\sigma_2)\Big)^2}{4\kappa_2\mu_2\Big(\kappa_1\mu_1(\mu_2+\kappa_2)+(1-\theta)\Big(\mu_1\mu_2(\kappa_2-\kappa_1)+\kappa_1\kappa_2(\mu_2-\mu_1)\Big)\Big)}.
\end{aligned}$$

This bound is achieved if and only if (3.31) satisfies $|\eta_1|<|\eta_2|$, i.e. if and only if

$$\begin{aligned}
&\Big((1-\theta)\delta\kappa\mu_2(\sigma_1+\sigma_2)+\Big(\kappa_1(\mu_2+\kappa_2)+(1-\theta)\delta\kappa\mu_2\Big)(\sigma_1-\sigma_2)\Big)\Big((1-\theta)\delta\mu\kappa_2(\sigma_1-\\
&\qquad-\sigma_2)+\Big(\mu_1(\mu_2+\kappa_2)+(1-\theta)\delta\mu\kappa_2\Big)(\sigma_1+\sigma_2)\Big)<0,
\end{aligned}$$

which is equivalent to the condition

$$\begin{cases}
(1-\theta)\delta\kappa\mu_2(\sigma_1+\sigma_2) < -\Big(\kappa_1(\mu_2+\kappa_2)+(1-\theta)\delta\kappa\mu_2\Big)(\sigma_1-\sigma_2),\\
-(1-\theta)\delta\mu\kappa_2(\sigma_1-\sigma_2) < \Big(\mu_1(\mu_2+\kappa_2)+(1-\theta)\delta\mu\kappa_2\Big)(\sigma_1+\sigma_2),
\end{cases} \tag{3.32}$$

(this happens only if $\sigma_1 + \sigma_2 \geq 0 \,\&\, \sigma_1 - \sigma_2 \leq 0$) or

$$\begin{cases} (1-\theta)\delta\kappa\mu_2(\sigma_1+\sigma_2) > -\Big(\kappa_1(\mu_2+\kappa_2) + (1-\theta)\delta\kappa\mu_2\Big)(\sigma_1-\sigma_2), \\ -(1-\theta)\delta\mu\kappa_2(\sigma_1-\sigma_2) > \Big(\mu_1(\mu_2+\kappa_2) + (1-\theta)\delta\mu\kappa_2\Big)(\sigma_1+\sigma_2), \end{cases} \tag{3.33}$$

(which happens only if $\sigma_1 + \sigma_2 \leq 0 \,\&\, \sigma_1 - \sigma_2 \geq 0$).

II. The case $|\eta_1| > |\eta_2|$ is *symmetric* to the previous one, and it is sufficient to change the positions of σ_1 and σ_2 in expressions for η_1 and η_2 as well as in conditions (3.32) and (3.33).

Note that conditions for the first and the second case can be jointly rewritten as

$$\begin{aligned} (1-\theta)\mu_2(\kappa_2-\kappa_1)|\sigma_1+\sigma_2| &< \Big(\kappa_1(\kappa_2+\mu_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)|\sigma_1-\sigma_2|, \\ (1-\theta)\kappa_2(\mu_2-\mu_1)|\sigma_1-\sigma_2| &< \Big(\mu_1(\kappa_2+\mu_2) + (1-\theta)\kappa_2(\mu_2-\mu_1)\Big)|\sigma_1+\sigma_2|, \end{aligned} \tag{3.34}$$

and the bound is then given with

$$\begin{aligned} \mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq{}& \theta\mathbf{A}_1^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + (1-\theta)\mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} - \\ &- (1-\theta)\theta\frac{(\kappa_2+\mu_2)\Big(\kappa_1\kappa_2(\mu_2-\mu_1)|\sigma_1-\sigma_2| + \mu_1\mu_2(\kappa_2-\kappa_1)|\sigma_1+\sigma_2|\Big)^2}{4\kappa_2\mu_2\Big(\kappa_1\mu_1(\mu_2+\kappa_2) + (1-\theta)\Big(\mu_1\mu_2(\kappa_2-\kappa_1) + \kappa_1\kappa_2(\mu_2-\mu_1)\Big)\Big)}, \end{aligned}$$

which proves part A of the theorem.

III. If the condition (3.34) is not satisfied, the maximum of the function $F(\eta_1, \eta_2)$ is attained on $|\eta_1| = |\eta_2|$, namely when $\eta_1 = \eta_2$, or $\eta_1 = -\eta_2$. In the first case, the maximum value

$$\max_{(\eta_1,\eta_2)\in\mathbf{R}^2} F(\eta_1,\eta_2) = \frac{\delta\kappa(\kappa_2+\mu_2)(\sigma_1+\sigma_2)^2}{4\kappa_2\Big(\kappa_1(\mu_2+\kappa_2) + (1-\theta)\mu_2\delta\kappa\Big)} \tag{3.35}$$

is attained for

$$\eta_1 = \eta_2 = \frac{\delta\kappa(\mu_2+\kappa_2)(\sigma_1+\sigma_2)}{4\kappa_2\Big(\kappa_1(\mu_2+\kappa_2) + (1-\theta)\mu_2\delta\kappa\Big)}, \tag{3.36}$$

while in the second one the maximum value

$$\max_{(\eta_1,\eta_2)\in\mathbf{R}^2} F(\eta_1,\eta_2) = \frac{\delta\mu(\kappa_2+\mu_2)(\sigma_1-\sigma_2)^2}{4\mu_2\Big(\mu_1(\mu_2+\kappa_2) + (1-\theta)\kappa_2\delta\mu\Big)} \tag{3.37}$$

is attained for

$$\eta_1 = -\eta_2 = \frac{\delta\mu(\mu_2+\kappa_2)(\sigma_1-\sigma_2)}{4\mu_2\Big(\mu_1(\mu_2+\kappa_2) + (1-\theta)\kappa_2\delta\mu\Big)}. \tag{3.38}$$

It can easily be showed that (3.35) is greater then (3.37) when the first inequality in (3.34) is not valid, and (3.35) is less then (3.37) when the second one is not valid. This implies the bound

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{\theta(\kappa_2 - \kappa_1)(\kappa_2 + \mu_2)(\sigma_1 + \sigma_2)^2}{4\kappa_2\left((1-\theta)\kappa_2\mu_2 + \kappa_1(\kappa_2 + \theta\mu_2)\right)}$$

when $(1-\theta)\mu_2(\kappa_2 - \kappa_1)|\sigma_1 + \sigma_2| \geq \left(\kappa_1(\kappa_2 + \mu_2) + (1-\theta)\mu_2(\kappa_2 - \kappa_1)\right)(\sigma_2 - \sigma_1)$ and the bound

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{\theta(\mu_2 - \mu_1)(\kappa_2 + \mu_2)(\sigma_1 - \sigma_2)^2}{4\mu_2\left((1-\theta)\kappa_2\mu_2 + \mu_1(\mu_2 + \theta\kappa_2)\right)}$$

when $(1-\theta)\kappa_2(\mu_2 - \mu_1)(\sigma_2 - \sigma_1) \geq \left(\mu_1(\kappa_2 + \mu_2) + (1-\theta)\kappa_2(\mu_2 - \mu_1)\right)|\sigma_1 + \sigma_2|$, which concludes the first part of the proof.

It remains to find the optimal microstructure for each case. As it was stated in Proposition 1.26, the optimal microstructure can be found among sequential laminates, with lamination directions given by the extremal vectors in

$$g^c(\boldsymbol{\eta}) = \max_{\mathsf{e}\in S^1}\left(f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta}\right), \tag{3.39}$$

which can be found among the eigenvectors of $\boldsymbol{\sigma}$. Recall the proof of Proposition 1.26, if the function g^c is differentiable in optimal $\boldsymbol{\eta}$, then the necessary condition of optimality for (3.28) reads

$$\boldsymbol{\sigma} - \left(\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1}\right)^{-1}\boldsymbol{\eta} = (1-\theta)f_2^c(\mathsf{e})\boldsymbol{\eta}, \tag{3.40}$$

where

$$f_2^c(\mathsf{e})\boldsymbol{\eta} = \mathbf{A}_2\boldsymbol{\eta} - \frac{1}{\mu_2}\left[\mathbf{A}_2\left((\mathbf{A}_2\boldsymbol{\eta})\mathsf{e}\otimes\mathsf{e} - ((\mathbf{A}_2\boldsymbol{\eta})\mathsf{e}\cdot\mathsf{e})\frac{\kappa_2}{\mu_2+\kappa_2}\mathsf{e}\otimes\mathsf{e}\right)\right].$$

and e is the extremal for (3.39). Otherwise, the optimality condition reads

$$\boldsymbol{\sigma} - \left(\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1}\right)^{-1}\boldsymbol{\eta} = (1-\theta)\sum_{i=1}^{p} m_i f_2^c(\mathsf{e}_i)\boldsymbol{\eta}, \tag{3.41}$$

where $m_i \geq 0$, $\sum_{i=1}^{p} m_i = 1$ and each e_i is extremal for (3.39).

Since the function g^c is differentiable everywhere, except on $|\eta_1| = |\eta_2|$, we shall distinguish three cases: $|\eta_1| \neq |\eta_2|$, $\eta_1 = \eta_2$ and $\eta_1 = -\eta_2$.

I. If $|\eta_1| < |\eta_2|$ or $|\eta_1| > |\eta_2|$, then the function $g^c(\boldsymbol{\eta})$ is differentiable, and let us show that the optimal microstructure is a simple laminate given by

$$\theta\left(\mathbf{A}^{-1} - \mathbf{A}_2^{-1}\right)^{-1} = \left(\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1}\right)^{-1} + (1-\theta)f_2^c(\mathsf{e}), \tag{3.42}$$

where the lamination direction e is extremal for (3.39). From Lemma 1.22 the lamination direction is the eigenvector associated to the eigenvalue of smaller absolute value of $\boldsymbol{\eta}$ (this is also an eigenvector of $\boldsymbol{\sigma}$, since it is simultaneously diagonalizable with $\boldsymbol{\eta}$). Multiplying (3.42) by $\boldsymbol{\eta}$, and using the necessary condition of optimality (3.40), it follows

$$\mathbf{A}^{-1}\boldsymbol{\sigma} = \mathbf{A}_2^{-1}\boldsymbol{\sigma} + \theta\boldsymbol{\eta}.$$

By taking the inner product with $\boldsymbol{\sigma}$ in the above equation, and the inner product with $\boldsymbol{\eta}$ in (3.40), a simple calculation leads to

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} = \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta\left(2\boldsymbol{\sigma} : \boldsymbol{\eta} - \left(\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1}\right)^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta) f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta}\right),$$

which proves that (3.42) is the optimal microstructure that corresponds to case A of the theorem.

II. Let $\eta_1 = \eta_2$, given by (3.36) and denote by e_1 and e_2 eigenvectors of $\boldsymbol{\sigma}$. These vectors are also eigenvectors for $\boldsymbol{\eta}$ and extremal for (3.39). By denoting $k = \frac{\delta\kappa(\mu_2+\kappa_2)(\sigma_1+\sigma_2)}{4\kappa_2(\kappa_1(\mu_2+\kappa_2)+(1-\theta)\mu_2\delta\kappa)}$, optimality condition (3.41) reduces to the system

$$\begin{aligned}
\sigma_1 - \frac{2\kappa_1\kappa_2}{\delta\kappa}k &= \frac{4\mu_2\kappa_2}{\mu_2+\kappa_2}k(1-\theta)m_2\\
\sigma_2 - \frac{2\kappa_1\kappa_2}{\delta\kappa}k &= \frac{4\mu_2\kappa_2}{\mu_2+\kappa_2}k(1-\theta)m_1,
\end{aligned}$$

from which we get parameters

$$\begin{aligned}
m_1 &= \frac{1}{2} + \frac{\Big(\kappa_1(\mu_2+\kappa_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)(\sigma_2-\sigma_1)}{2(1-\theta)(\kappa_2-\kappa_1)\mu_2(\sigma_1+\sigma_2)}\\
m_2 &= \frac{1}{2} + \frac{\Big(\kappa_1(\mu_2+\kappa_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)(\sigma_1-\sigma_2)}{2(1-\theta)(\kappa_2-\kappa_1)\mu_2(\sigma_1+\sigma_2)}.
\end{aligned}$$

It is straightforward that $m_1 + m_2 = 1$, while their nonnegativity follows from the condition

$$(1-\theta)\mu_2(\kappa_2-\kappa_1)|\sigma_1+\sigma_2| \geq \Big(\kappa_1(\kappa_2+\mu_2) + (1-\theta)\mu_2(\kappa_2-\kappa_1)\Big)|\sigma_1-\sigma_2|,$$

which determines case B of the theorem. The bound (3.23) is achieved by the rank-2 sequential laminate with lamination direction e_1 and e_2 and lamination parameters m_1 and m_2 given by (3.24). This can be seen by using analogous procedure as in case I.

III. Finally, assume that $\eta_1 = -\eta_2$, given by (3.38) and denote by e_1 and e_2 eigenvectors of $\boldsymbol{\sigma}$. Again, these vectors are also eigenvectors of $\boldsymbol{\eta}$ and extremal for (3.39), by

Lemma 1.22. By denoting $l = \frac{\delta\mu(\mu_2+\kappa_2)(\sigma_1-\sigma_2)}{4\mu_2(\mu_1(\mu_2+\kappa_2)+(1-\theta)\kappa_2\delta\mu)}$, from the optimality condition (3.41) we obtain

$$
\begin{aligned}
\sigma_1 - \frac{2\mu_1\mu_2}{\delta\mu} l &= \frac{4\mu_2\kappa_2}{\mu_2+\kappa_2} l(1-\theta)m_2 \\
\sigma_2 + \frac{2\mu_1\mu_2}{\delta\mu} l &= -\frac{4\mu_2\kappa_2}{\mu_2+\kappa_2} l(1-\theta)m_1,
\end{aligned}
$$

and thus

$$
\begin{aligned}
m_1 &= \frac{1}{2} + \frac{\Big(\mu_1(\mu_2+\kappa_2)+(1-\theta)\kappa_2(\mu_2-\mu_1)\Big)(\sigma_1+\sigma_2)}{2(1-\theta)(\mu_2-\mu_1)\kappa_2(\sigma_2-\sigma_1)} \\
m_2 &= \frac{1}{2} + \frac{\Big(\mu_1(\mu_2+\kappa_2)+(1-\theta)\kappa_2(\mu_2-\mu_1)\Big)(\sigma_1+\sigma_2)}{2(1-\theta)(\mu_2-\mu_1)\kappa_2(\sigma_1-\sigma_2)}.
\end{aligned}
$$

Obviously, $m_1 + m_2 = 1$, and from the inequality that determines case C of the theorem, it follows that m_1 and m_2 are nonnegative. The bound (3.26) is achieved by the rank-2 sequential laminate with lamination directions e_1 and e_2 and lamination parameters m_1 and m_2 given by (3.27), which concludes the proof.

□

3.1.2 Numerical examples

In this section we shall present numerical solutions of some compliance minimization problems, applying Algorithm 3.4. The state equation is solved by the finite element method in the deal.II library (Bangerth et al. (2000)) using Lagrange elements on a fine mesh. Moreover, a design $(\theta, \mathbf{A})$ is discretized on the same mesh, by piecewise constant elements. The Lagrange multiplier is recalculated in each step in order to satisfy the volume constraint. We shall present the result of the 20th iteration of the algorithm, although a similar design is reached already by its first few iterations.

Example 3.1 (Compliance minimization on a ball.) In the first example we take a ball $B(\mathbf{0}, 2)$ as domain Ω, filled with isotropic phases $\mathbf{A}_1$ and $\mathbf{A}_2$ with shear moduli $\mu_1 = 63$, $\mu_2 = 75$, and bulk moduli $\kappa_1 = 58$, $\kappa_2 = 139$, and with a volume constraint of 25% for the first material. We consider the compliance minimization

$$
J(\theta, \mathbf{A}) = \int_\Omega \mathsf{f}(\mathbf{x}) \cdot \mathsf{u}(\mathbf{x})\, d\mathbf{x} + l \int_\Omega \theta(\mathbf{x})\, d\mathbf{x} \longrightarrow \min,
$$

where the function u is the solution of the linearized elasticity system

$$
\begin{cases}
-\mathsf{div}\,(\mathbf{A}e(\mathsf{u})) = \mathsf{f} \quad \text{in } \Omega \\
\mathsf{u} \in \mathrm{H}^1_0(\Omega; \mathbf{R}^d)\,,
\end{cases}
$$

with the force term $\mathsf{f} = 100\mathsf{e}_r$, where e_r is the unit vector in radial direction. For the initial design we take $\theta^0 = 0.25$ and a simple laminate $\mathbf{A}^0$ with e_1 as the lamination direction. The numerical solution is shown in Figure 3.1. The red part of the domain corresponds to phase $\mathbf{A}_1$, while the yellow one to the phase $\mathbf{A}_2$. The orange tones corresponds to the composite material. Convergence history is presented in Figure 3.2.

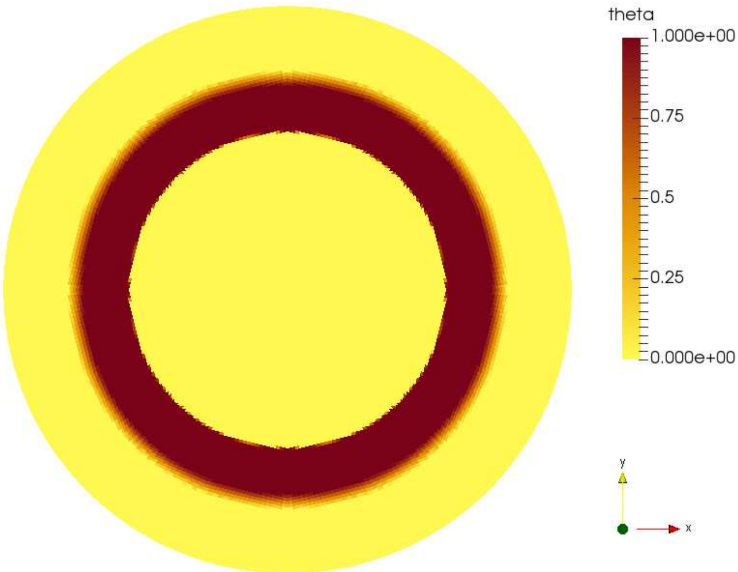
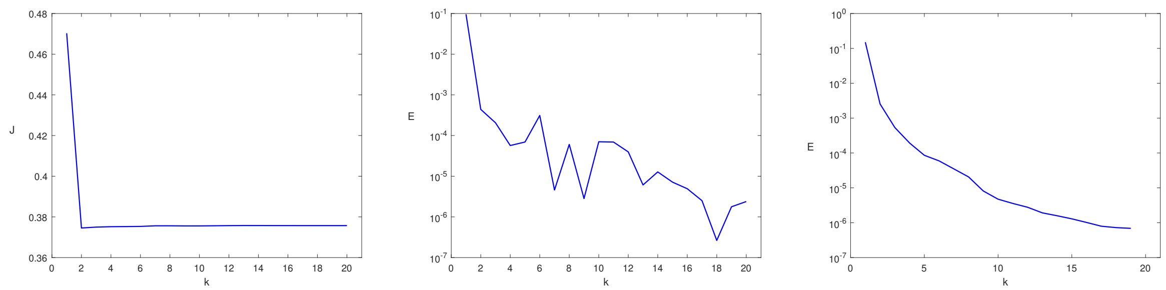


Figure 3.1: Numerical solution for the optimal design problem presented in Example 3.1.



(a) Cost functional J in terms of the iteration number k.

(b) $E = |J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ in terms of the iteration number k.

(c) $E = \|\theta^k - \theta^{k+1}\|^2_{L^2}$ in terms of the iteration number k.

Figure 3.2: Convergence history for the optimal design problem presented in Example 3.1.

Example 3.2 (Compliance minimization on a rectangle.) For the second example we take $\Omega = [-2, 2] \times [0, 1]$ and consider compliance minimization

$$\int_\Omega \mathsf{f} \cdot \mathsf{u} \, d\mathbf{x} + \int_{\Gamma_N} \mathsf{g} \cdot \mathsf{u} \, dS \longrightarrow \min,$$

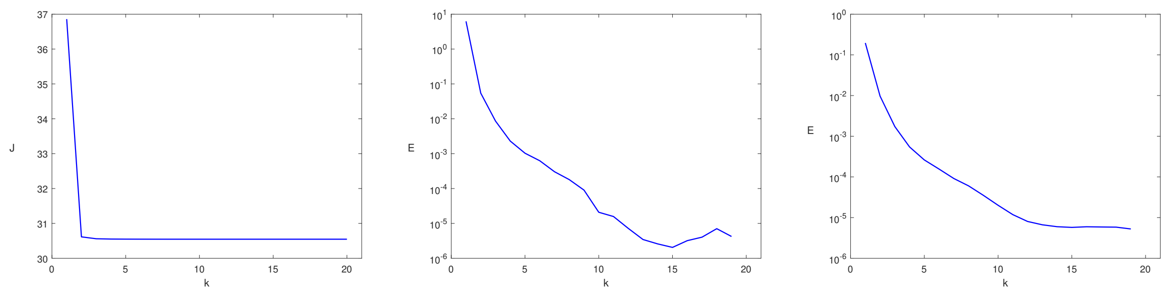
where u is the solution of the linearized elasticity system with mixed boundary conditions,

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}e(\mathsf{u})) = \mathsf{f} & \text{in } \Omega \\ \mathsf{u} = \mathbf{0} & \text{on } \Gamma_D \\ \mathbf{A}e(\mathsf{u})\mathsf{n} = \mathsf{g} & \text{on } \Gamma_N. \end{cases} \tag{3.43}$$

The boundary part Γ_D corresponds to $\partial\Omega \cap (B((-2,0),0.1) \cup B((2,0),0.1))$, while Γ_N corresponds to the rest of the boundary. We take materials as in Example 3.1, $\mathsf{f} = -10 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathsf{g} = \mathbf{0}$, and 50% as the overall portion of the first phase.

A numerical solution to this optimal design problem is presented in Figure 3.3. Again, the red part of the domain corresponds to weaker phase $\mathbf{A}$, the yellow part of the domain to the stronger phase $\mathbf{A}_2$, and between them a composite material occurs. The convergence history is shown in Figure 3.4.

Figure 3.3: Numerical solution for the optimal design problem presented in Example 3.2.



(a) Cost functional J in terms of the iteration number k.

(b) $E = |J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ in terms of the iteration number k.

(c) $E = \|\theta^k - \theta^{k+1}\|_{L^2}^2$ in terms of the iteration number k.

Figure 3.4: Convergence history for the optimal design problem presented in Example 3.2.

Example 3.3 (Compliance minimization on a rectangle.) In the third example the domain is $\Omega = [0,1] \times [0,3]$ and we consider a similar problem as in the previous example, i.e compliance minimization

$$\int_\Omega \mathsf{f} \cdot \mathsf{u} \, d\mathbf{x} + \int_{\Gamma_N} \mathsf{g} \cdot \mathsf{u} \, dS \longrightarrow \min,$$

where u is the solution of the linearized elasticity system with mixed boundary conditions,

$$\begin{cases} -\mathsf{div}\,(\mathbf{A}e(\mathsf{u})) = \mathbf{0} & \text{in } \Omega \\ \mathsf{u} = \mathbf{0} & \text{on } \Gamma_D \\ \mathbf{A}e(\mathsf{u})\mathsf{n} = \mathsf{g} & \text{on } \Gamma_N. \end{cases} \tag{3.44}$$

The boundary part Γ_D corresponds to the left part of $\partial\Omega$, where $x = 0$, while Γ_N corresponds to the rest of the boundary. We take the first phase with elastic moduli $\mu_1 = 13$, $\kappa_1 = 7$, and the second one with ten times stronger moduli. Moreover, we set function $\mathsf{g} = -10 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \chi_{B((1,1.5),0.02)}$, and take 90% as the overall portion of the first phase.

A numerical solution for the above optimal design problem is presented in Figure 3.5, while the convergence history is shown in Figure 3.6.

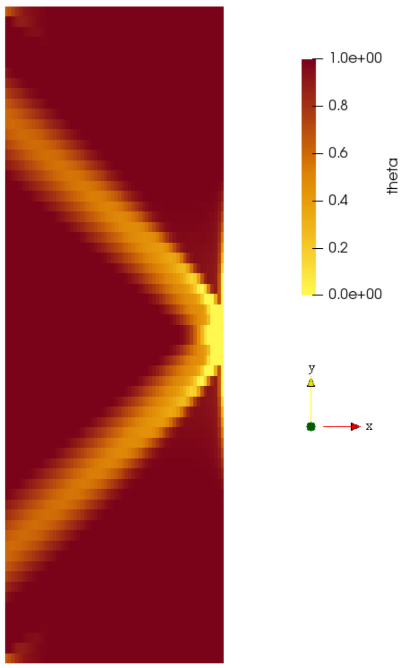
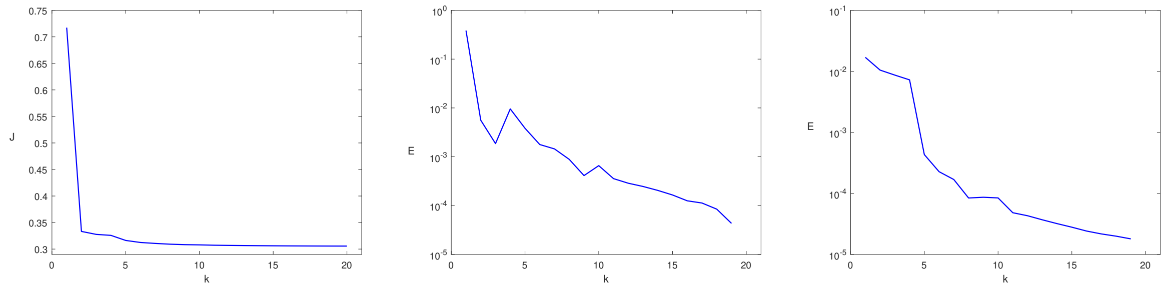


Figure 3.5: Numerical solution for the optimal design problem presented in Example 3.3.



(a) Cost functional J in terms of the iteration number k.

(b) $E = |J(\theta^k, \mathbf{A}^k) - J(\theta^{k+1}, \mathbf{A}^{k+1})|$ in terms of the iteration number k.

(c) $E = \|\theta^k - \theta^{k+1}\|^2_{L^2}$ in terms of the iteration number k.

Figure 3.6: Convergence history for the optimal design problem presented in Example 3.3.

3.2 On Hashin-Shtrikman bounds in 3D linearized elasticity

As it was seen in Section 3.1, in order to develop an optimality criteria method for optimal design problems in linearized elasticity, explicit calculation of the lower Hashin-Shtrikman bound on the complementary energy is needed, as well as the optimal microsturucture. In the two-dimensional case it was done quite simply in Theorem 3.5. In the three-dimensional case, the algebra becomes formidable, thus, we shall use Mathematica software for some calculations. From Proposition 1.21 if $d = 3$, for any $\boldsymbol{\sigma} \in \mathrm{Sym}_3$ the lower Hashin-Shtrikman bound reads

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta \max_{\boldsymbol{\eta}\in\mathrm{Sym}_3} \Big[2\boldsymbol{\sigma} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta)g^c(\boldsymbol{\eta})\Big], \quad (3.45)$$

where $\mathbf{A}_1$ and $\mathbf{A}_2$ are isotropic elastic phases with bulk moduli κ_1 and κ_2, and shear moduli μ_1 and μ_2 respectively. Function $g^c(\boldsymbol{\eta}) = \max_{\mathsf{e}\in S^2}(f_2^c(\mathsf{e})\boldsymbol{\eta} : \boldsymbol{\eta})$ is explicitly given in Lemma 1.22, and for eigenvalues $\eta_1 \leq \eta_2 \leq \eta_3$ of the matrix $\boldsymbol{\eta}$, it reads

$$g^c(\boldsymbol{\eta}) = \mathbf{A}_2\boldsymbol{\eta} : \boldsymbol{\eta} - \frac{1}{2\mu_2 + \lambda_2}\min\{(2\mu_2\eta_1 + \lambda_2\mathrm{tr}\boldsymbol{\eta})^2, (2\mu_2\eta_3 + \lambda_2\mathrm{tr}\boldsymbol{\eta})^2\},$$

where $\lambda_2 := \kappa_2 - \frac{2\mu_2}{3}$. Before stating the theorem, let us introduce some notation. We denote

$$\begin{aligned}
&\delta\kappa := \kappa_2 - \kappa_1,\\
&\delta\mu := \mu_2 - \mu_1,\\
&\zeta_1 := 9\kappa_2\mu_2\Big(4\theta\delta\kappa\delta\mu(3\kappa_2 + \mu_2)^2 + \kappa_2\mu_2\Big(3\kappa_1(12\kappa_2 + 9\mu_1 + 7\mu_2) - 4(9{\kappa_2}^2 +\\
&\quad + 6\kappa_2(\mu_1 + \mu_2) + \mu_2\delta\mu)\Big)\Big),
\end{aligned}$$

$$\zeta_2 := 36\kappa_2\mu_2\Big(\mu_2(\kappa_2-\theta\delta\kappa)(\mu_1+3\mu_2-3\theta\delta\mu)+3\kappa_1\kappa_2(\mu_2-\theta\delta\mu)\Big),$$
$$\zeta_3 := 36\kappa_2\mu_2\Big(3\kappa_1\kappa_2(\theta\delta\mu-\mu_2)+4\mu_1\mu_2(\theta\delta\kappa-\kappa_2)\Big),$$
$$\zeta_4 := 36\kappa_2\mu_2\Big(3\kappa_2(\mu_2-\theta\delta\mu)+\mu_2(3\mu_1+\mu_2+\theta\delta\mu)\Big)\Big(\theta\delta\kappa\delta\mu(3\kappa_2+\mu_2)-\kappa_2\mu_2(3\delta\kappa+\delta\mu)\Big),$$

and define functions $f_1, f_2, f_3, f_4 : \mathbf{R}^3 \to \mathbf{R}$ with

$$\begin{aligned} f_1(x,y,z) = 4\kappa_2\mu_2 \Big(&3\kappa_2 \Big(\mu_1 \left((5-3\theta)z(x+y)+(3\theta-2)((x+y)^2+z^2)\right)+\\ &+3(1-\theta)\mu_2(-z(x+y)+(x+y)^2+z^2)\Big)+\mu_2(3(1-\theta)\delta\mu+4\mu_1)(x+y+z)^2\Big)-\\ &-\kappa_1\Big(12\kappa_2\mu_2(z(x+y)((1-3\theta)\delta\mu+2\mu_1)+(x+y)^2((1-3\theta)\delta\mu+2\mu_2)+\\ &+z^2(\mu_1-(1+3\theta)\delta\mu))+4\mu_2^2(3(1-\theta)\delta\mu+4\mu_1)(x+y+z)^2-9\kappa_2^2\delta\mu(x+y-2z)^2\Big),\end{aligned}$$

$$\begin{aligned} f_2(x,y,z) = \kappa_1\Big(&-12\kappa_2\mu_2\left(\mu_1\left((4-9\theta)x^2+2x(y+z)-2(y+z)^2\right)+\mu_2\left((9\theta-5)x^2-\right.\right.\\ &\left.\left.-4x(y+z)+(y+z)^2\right)\right)-9\kappa_2^2\delta\mu(-2x+y+z)^2+16\mu_1\mu_2^2(x+y+z)^2\Big)-\\ &-4\kappa_2\mu_2\Big(3\kappa_2(\mu_1((9\theta-8)x^2+2x(y+z)+(y+z)^2)+9(1-\theta)\mu_2x^2)+\\ &+4\mu_1\mu_2(x+y+z)^2\Big),\end{aligned}$$

$$\begin{aligned} f_3(x,y,z) = 27\theta^2\kappa_2\mu_2\delta\kappa\delta\mu^2x^2+\theta\delta\mu\Big(&\kappa_1\Big(-9\kappa_2^2\delta\mu\left(x^2-x(y+z)+y^2-yz+z^2\right)-\\ &-3\kappa_2\mu_2\Big(\mu_1(4x^2+2x(y+z)-5y^2+2yz-5z^2)+\mu_2\Big(-14x^2-4x(y+z)+\\ &+(y+z)^2\Big)\Big)+4\mu_1\mu_2^2(x+y+z)^2\Big)-2\kappa_2\mu_2\Big(3\kappa_2\Big(\mu_1\Big(-4x^2+x(y+z)+2(y^2-\\ &-yz+z^2)\Big)+9\mu_2x^2\Big)+2\mu_1\mu_2(x+y+z)^2\Big)\Big)+\mu_2\Big(\kappa_1\Big(9\kappa_2^2\delta\mu(x^2-x(y+z)+y^2-\\ &-yz+z^2)+3\kappa_2\mu_2(x+y+z)(4\mu_1x-2\mu_1(y+z)+\mu_2(-5x+y+z))-4\mu_1\mu_2^2(x+\\ &+y+z)^2\Big)+\kappa_2\Big(3\kappa_2\Big(-3\mu_1^2(y-z)^2+2\mu_1\mu_2\Big(-4x^2+x(y+z)+2(y^2-yz+z^2)\Big)+\\ &+9\mu_2^2x^2\Big)+4\mu_1\mu_2^2(x+y+z)^2\Big)\Big),\end{aligned}$$

and

$$\begin{aligned} f_4(x,y,z) = \delta\mu\Big(&4\theta\delta\kappa\delta\mu(3\kappa_2+\mu_2)\Big(9\kappa_2^2(x^2-x(y+z)+y^2-yz+z^2)+3\kappa_2\mu_2(2x^2+\\ &+x(y-8z)+2y^2+yz+2z^2)+\mu_2^2(x+y+z)^2\Big)+\mu_2\Big(\kappa_1\Big(108\kappa_2^3\Big(x^2-x(y+z)+\\ &+y^2-yz+z^2\Big)+27\kappa_2^2\Big(\mu_1(x-2y+z)^2+\mu_2(3x-z)(x-3z)+4\mu_2y^2\Big)+\\ &+36\kappa_2\mu_2(x+y+z)\Big(\mu_2y-\mu_1(x-2y+z)\Big)+4\mu_2^2(3\mu_1+\mu_2)(x+y+z)^2\Big)-\\ &-4\kappa_2\Big(27\kappa_2^3\Big(x^2-x(y+z)+y^2-yz+z^2\Big)+27\kappa_2^2\Big(\mu_2(x^2-3xz+y^2+z^2)-\\ &-\mu_1(x-y)(y-z)\Big)+9\kappa_2\mu_2\Big(\mu_1(-2x^2+xy+2y^2+yz-2z^2)+\mu_2(x^2+xy-\\ &-2xz+y^2+yz+z^2)\Big)+\mu_2^2(3\mu_1+\mu_2)(x+y+z)^2\Big)\Big)\Big).\end{aligned}$$

Theorem 3.6 Let $d = 3$ and $\theta \in [0,1]$, $0 < \mu_1 < \mu_2$, $0 < \kappa_1 < \kappa_2$ given. For eigenvalues $\sigma_1 \leq \sigma_2 \leq \sigma_3$ of the matrix $\boldsymbol{\sigma}$, the lower Hashin-Shtrikman bound on complementary energy (3.45) is explicitly given as follows:

A. If

$$\begin{aligned} 3\kappa_2(\sigma_1 + \sigma_2 - 2\sigma_3) + \mu_2(\sigma_1 + \sigma_2 + \sigma_3) \geq 0, \\ 3\kappa_2(2\sigma_1 - \sigma_2 - \sigma_3) - \mu_2(\sigma_1 + \sigma_2 + \sigma_3) \geq 0, \end{aligned} \tag{3.46}$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma}. \tag{3.47}$$

B. If

$$\begin{aligned} \sigma_1 + \sigma_2 + \sigma_3 \geq 0, \\ 12(1-\theta)\delta\kappa\mu_2\sigma_1 - \kappa_1(3\kappa_2 + 4\mu_2)(-2\sigma_1 + \sigma_2 + \sigma_3) \geq 0, \\ \Big(\kappa_1(3\kappa_2 - 2\mu_2) + 6\mu_2(\kappa_2 - \theta\delta\kappa)\Big)(\sigma_1 + \sigma_2 - \sigma_3) - \kappa_1(3\kappa_2 + 4\mu_2)\sigma_3 \geq 0, \end{aligned} \tag{3.48}$$

or

$$\begin{aligned} \sigma_1 + \sigma_2 + \sigma_3 \leq 0, \\ 12(1-\theta)\delta\kappa\mu_2\sigma_3 - \kappa_1(3\kappa_2 + 4\mu_2)(\sigma_1 + \sigma_2 - 2\sigma_3) \leq 0, \\ \Big(\kappa_1(3\kappa_2 - 2\mu_2) + 6\mu_2(\kappa_2 - \theta\delta\kappa)\Big)(-\sigma_1 + \sigma_2 + \sigma_3) - \kappa_1(3\kappa_2 + 4\mu_2)\sigma_1 \leq 0, \end{aligned} \tag{3.49}$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \theta\frac{\delta\kappa(3\kappa_2 + 4\mu_2)(\sigma_1 + \sigma_2 + \sigma_3)^2}{9\kappa_2\Big(4(1-\theta)\kappa_2\mu_2 + \kappa_1(3\kappa_2 + 4\theta\mu_2)\Big)}. \tag{3.50}$$

C. If

$$\begin{aligned} 3\kappa_2(\sigma_1 + \sigma_2 - 2\sigma_3) + \mu_2(\sigma_1 + \sigma_2 + \sigma_3) \leq 0, \\ -2\theta\delta\kappa\delta\mu(3\kappa_2 + \mu_2)\Big((3\kappa_2 + \mu_2)(\sigma_1 - \sigma_2) - 3\mu_2\sigma_3\Big) + \mu_2\Big(2\kappa_2\Big(9\kappa_2^2(\sigma_1 - \\ -\sigma_2) + \mu_2\Big(\mu_2(\sigma_1 - \sigma_2 - 3\sigma_3) - \mu_1(5\sigma_1 + 3\sigma_2 + \sigma_3)\Big) + 3\kappa_2\Big(\mu_2(2\sigma_1 - \\ -2\sigma_2 - 3\sigma_3) + \mu_1(\sigma_1 - 3\sigma_2 + 2\sigma_3)\Big)\Big) + \kappa_1\Big(18\kappa_2^2(-\sigma_1 + \sigma_2) + \\ +2\mu_2(3\mu_1 + \mu_2)(\sigma_1 + \sigma_2 + \sigma_3) + 3\kappa_2\Big(-3\mu_1(\sigma_1 - 2\sigma_2 + \sigma_3) + \\ +\mu_2(-3\sigma_1 + 4\sigma_2 + 5\sigma_3)\Big)\Big)\Big) \geq 0, \end{aligned} \tag{3.51}$$

and either

$$
\begin{aligned}
-2\kappa_2\Big((-3+9\theta)\kappa_2\mu_1 + 9(1-\theta)\kappa_2\mu_2 + (5+3\theta)\mu_1\mu_2 + 3(1- \\
-\theta)\mu_2^2\Big)(\sigma_1+\sigma_2) + \kappa_1\Big(3\kappa_2\Big((-3+6\theta)\mu_1 + (7-6\theta)\mu_2\Big) + 2\mu_2\Big(3(1+ \\
+\theta)\mu_1 + (5-3\theta)\mu_2\Big)\Big)(\sigma_1+\sigma_2) + 2\kappa_1\Big(\kappa_2(9\mu_1 - 3\mu_2) + \mu_2\Big((3+ \\
+9\theta)\mu_1 + (5-9\theta)\mu_2\Big)\Big)\sigma_3 - 2\kappa_2\Big(6\kappa_2\mu_1 - \mu_2\Big(\mu_1 - 9\theta\mu_1 - \\
-9(1-\theta)\mu_2\Big)\Big)\sigma_3 \geq 0,
\end{aligned}
\tag{3.52}
$$

or

$$
\begin{aligned}
(3\kappa_2+4\mu_2)\Big(-4\kappa_2\mu_1 + \kappa_1(3\mu_1+\mu_2)\Big)(\sigma_1+\sigma_2) + 2\Big(3\kappa_2\Big((-3+ \\
+6\theta)\kappa_1 + (4-6\theta)\kappa_2\Big)\mu_1 + \Big(3\kappa_2\Big((5-6\theta)\kappa_1 + 6(-1+\theta)\kappa_2\Big) - \\
-2\Big(-3(1+\theta)\kappa_1 + \kappa_2 + 3\theta\kappa_2\Big)\mu_1\Big)\mu_2 + 2\Big(\kappa_1 - 3\theta\kappa_1 - \\
-3(1-\theta)\kappa_2\Big)\mu_2^2\Big)\sigma_3 \leq 0,
\end{aligned}
\tag{3.53}
$$

then

$$
\begin{aligned}
&\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} - \\
&\quad - \frac{\theta}{\zeta_1}\delta\kappa\delta\mu(3\kappa_2+4\mu_2)\Big(3\kappa_2(\sigma_1+\sigma_2-2\sigma_3) + \mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)^2.
\end{aligned}
\tag{3.54}
$$

D. If

$$
\begin{aligned}
\Big(\kappa_1(3\kappa_2 - 2\mu_2) + 6\mu_2(\kappa_2 - \theta\delta\kappa)\Big)(\sigma_1+\sigma_2-\sigma_3) - \kappa_1(3\kappa_2+4\mu_2)\sigma_3 \leq 0 \\
-2\kappa_2\Big((-3+9\theta)\kappa_2\mu_1 + 9(1-\theta)\kappa_2\mu_2 + (5+3\theta)\mu_1\mu_2 + 3(1- \\
-\theta)\mu_2^2\Big)(\sigma_1+\sigma_2) + \kappa_1\Big(3\kappa_2\Big((-3+6\theta)\mu_1 + (7-6\theta)\mu_2\Big) + 2\mu_2\Big(3(1+ \\
+\theta)\mu_1 + (5-3\theta)\mu_2\Big)\Big)(\sigma_1+\sigma_2) + 2\kappa_1\Big(\kappa_2(9\mu_1 - 3\mu_2) + \mu_2\Big((3+ \\
+9\theta)\mu_1 + (5-9\theta)\mu_2\Big)\Big)\sigma_3 - 2\kappa_2\Big(6\kappa_2\mu_1 + \mu_2\Big(9(1-\theta)\delta\mu + 8\mu_1\Big)\Big)\sigma_3 \leq 0 \\
-6\theta^2\delta\kappa\delta\mu\mu_2\sigma_1 + \mu_2\Big(-6\kappa_2\mu_2\sigma_1 + 3\kappa_1\kappa_2(-\sigma_1+\sigma_2) + 2\kappa_2\mu_1(\sigma_2-\sigma_3) \\
+2\kappa_1\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big) + \theta\Big(2\kappa_2\mu_2\Big(6\mu_2\sigma_1 + \mu_1(-3\sigma_1-\sigma_2+\sigma_3)\Big) \\
+\kappa_1\Big(3\kappa_2\delta\mu(\sigma_1-\sigma_2) + 2\mu_1\mu_2(\sigma_1+2\sigma_2) - 2\mu_2^2(4\sigma_1+\sigma_2+\sigma_3)\Big)\Big) \leq 0,
\end{aligned}
\tag{3.55}
$$

then

$$
\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{\theta}{\zeta_2} f_1(\sigma_1,\sigma_2,\sigma_3). \tag{3.56}
$$

E. If

$$
\begin{aligned}
&3\kappa_2(2\sigma_1-\sigma_2-\sigma_3)-\mu_2(\sigma_1+\sigma_2+\sigma_3)\leq 0\\
&-2\kappa_2\Big(9\kappa_2^2(-\theta\delta\mu+\mu_2)(\sigma_2-\sigma_3)+\mu_2^2\Big((1-\theta)\mu_2(3\sigma_1+\sigma_2-\sigma_3)+\\
&+\mu_1\Big((1+3\theta)\sigma_1+(3+\theta)\sigma_2-(-5+\theta)\sigma_3\Big)\Big)-3\kappa_2\mu_2\Big((-1+\\
&+\theta)\mu_2(3\sigma_1+2\sigma_2-2\sigma_3)+\mu_1\Big((2-3\theta)\sigma_1-(3+2\theta)\sigma_2+\\
&+(1+2\theta)\sigma_3\Big)\Big)\Big)+\kappa_1\Big(18\kappa_2^2(-\theta\delta\mu+\mu_2)(\sigma_2-\sigma_3)+2\mu_2^2\Big(\mu_1\Big(3(1+\\
&+\theta)(\sigma_1+\sigma_2+\sigma_3)-2\theta(\sigma_2+2\sigma_3)\Big)+\mu_2\Big((1-\theta)(\sigma_1+\sigma_2+\sigma_3)-\\
&-2\theta(\sigma_1-\sigma_3)\Big)\Big)+3\kappa_2\mu_2\Big(\mu_1\Big((-3+6\theta)\sigma_1+(6+4\theta)\sigma_2-(3+\\
&+4\theta)\sigma_3\Big)+\mu_2\Big((5-6\theta)\sigma_1+4(1-\theta)\sigma_2-(3-4\theta)\sigma_3\Big)\Big)\Big)\leq 0,
\end{aligned}
\tag{3.57}
$$

and either

$$
\begin{aligned}
&-4\kappa_2\Big(9(1-\theta)\kappa_2\mu_2\sigma_1+3(1-\theta)\mu_2^2\sigma_1+3\kappa_2\mu_1\Big((-2+3\theta)\sigma_1+\\
&+\sigma_2+\sigma_3\Big)+\mu_1\mu_2\Big((1+3\theta)\sigma_1+4(\sigma_2+\sigma_3)\Big)\Big)+\kappa_1\Big(4\mu_2\Big(\mu_2(\sigma_1-3\theta\sigma_1+\\
&+\sigma_2+\sigma_3)+3\mu_1(\sigma_1+\theta\sigma_1+\sigma_2+\sigma_3)\Big)+3\kappa_2\Big(\mu_2\Big(2(5-6\theta)\sigma_1+\\
&+\sigma_2+\sigma_3\Big)+3\mu_1\Big((-2+4\theta)\sigma_1+\sigma_2+\sigma_3\Big)\Big)\Big)\geq 0,
\end{aligned}
\tag{3.58}
$$

or

$$
\begin{aligned}
&2\kappa_2\Big(-9(1-\theta)\kappa_2\mu_2(\sigma_2+\sigma_3)-3(1-\theta)\mu_2^2(3\sigma_1+\sigma_2+\sigma_3)+\\
&+3\kappa_2\mu_1\Big(-2\sigma_1+(1-3\theta)(\sigma_2+\sigma_3)\Big)-\mu_1\mu_2\Big((-1+9\theta)\sigma_1+\\
&+(5+3\theta)(\sigma_2+\sigma_3)\Big)\Big)+\kappa_1\Big(3\kappa_2\Big(6\mu_1\sigma_1-2\mu_2\sigma_1+3(-1+2\theta)\mu_1(\sigma_2+\sigma_3)+\\
&+(7-6\theta)\mu_2(\sigma_2+\sigma_3)\Big)+2\mu_2\Big(3\mu_1\Big(\sigma_1+3\theta\sigma_1+(1+\theta)(\sigma_2+\sigma_3)\Big)+\\
&+\mu_2\Big((5-9\theta)\sigma_1+(5-3\theta)(\sigma_2+\sigma_3)\Big)\Big)\Big)\leq 0,
\end{aligned}
\tag{3.59}
$$

then

$$
\begin{aligned}
&\mathbf{A}^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}\geq\mathbf{A}_2^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}-\\
&-\frac{\theta}{\zeta_1}\delta\kappa\delta\mu(3\kappa_2+4\mu_2)\Big(3\kappa_2(-2\sigma_1+\sigma_2+\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)^2.
\end{aligned}
\tag{3.60}
$$

F. If

$$
\begin{aligned}
\sigma_2 - \sigma_3 \geq 0 \\
-12(1-\theta)\kappa_2\mu_2\sigma_1 + \kappa_1\Big(3\kappa_2(-2\sigma_1 + \sigma_2 + \sigma_3) + \\
+4\mu_2(\sigma_1 - 3\theta\sigma_1 + \sigma_2 + \sigma_3)\Big) \geq 0 \\
-4\kappa_2\Big(9(1-\theta)\kappa_2\mu_2\sigma_1 + 3(1-\theta)\mu_2^2\sigma_1 + 3\kappa_2\mu_1\Big((-2+3\theta)\sigma_1 + \\
\sigma_2 + \sigma_3\Big) + \mu_1\mu_2\Big((1+3\theta)\sigma_1 + 4(\sigma_2+\sigma_3)\Big)\Big) + \kappa_1\Big(4\mu_2\Big(\mu_2\Big((1-3\theta)\sigma_1 + \\
+\sigma_2 + \sigma_3\Big) + 3\mu_1(\sigma_1 + \theta\sigma_1 + \sigma_2 + \sigma_3)\Big) + 3\kappa_2\Big(\mu_2\Big(2(5-6\theta)\sigma_1 + \\
+\sigma_2 + \sigma_3\Big) + 3\mu_1(-2\sigma_1 + 4\theta\sigma_1 + \sigma_2 + \sigma_3)\Big)\Big) \leq 0.
\end{aligned}
\tag{3.61}
$$

then

$$
\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{\theta}{\zeta_3} f_2(\sigma_1, \sigma_2, \sigma_3). \tag{3.62}
$$

G. If

$$
\begin{aligned}
-2\theta\delta\kappa\delta\mu(3\kappa_2 + \mu_2)\Big((3\kappa_2 + \mu_2)(\sigma_1 - \sigma_2) - 3\mu_2\sigma_3\Big) + \\
+\mu_2\Big(2\kappa_2\Big(9\kappa_2^2(\sigma_1 - \sigma_2) + \mu_2\Big(\mu_2(\sigma_1 - \sigma_2 - 3\sigma_3) - \mu_1(5\sigma_1 + \\
+3\sigma_2 + \sigma_3)\Big) + 3\kappa_2\Big(\mu_2(2\sigma_1 - 2\sigma_2 - 3\sigma_3) + \mu_1(\sigma_1 - 3\sigma_2 + 2\sigma_3)\Big)\Big) + \\
+\kappa_1\Big(18\kappa_2^2(-\sigma_1 + \sigma_2) + 2\mu_2(3\mu_1 + \mu_2)(\sigma_1 + \sigma_2 + \sigma_3) + \\
+3\kappa_2\Big(- 3\mu_1(\sigma_1 - 2\sigma_2 + \sigma_3) + +\mu_2(-3\sigma_1 + 4\sigma_2 + 5\sigma_3)\Big)\Big)\Big) \leq 0, \\
-2\kappa_2\Big(9\kappa_2^2(-\theta\delta\mu + \mu_2)(\sigma_2 - \sigma_3) + \mu_2^2\Big((1-\theta)\mu_2(3\sigma_1 + \sigma_2 - \sigma_3) + \\
+\mu_1\Big((1+3\theta)\sigma_1 + (3+\theta)\sigma_2 - (-5+\theta)\sigma_3\Big)\Big) - 3\kappa_2\mu_2\Big(- (1- \\
-\theta)\mu_2(3\sigma_1 + 2\sigma_2 - 2\sigma_3) + \mu_1(2\sigma_1 - 3\theta\sigma_1 - 3\sigma_2 - 2\theta\sigma_2 + \sigma_3 + \\
+2\theta\sigma_3)\Big)\Big) + \kappa_1\Big(18\kappa_2^2(-\theta\delta\mu + \mu_2)(\sigma_2 - \sigma_3) + 2\mu_2^2\Big(\mu_1\Big(3(1+\theta)\sigma_1 + \\
+(3+\theta)\sigma_2 + +(3-\theta)\sigma_3\Big) + \mu_2(\sigma_1 - 3\theta\sigma_1 + \sigma_2 - \theta\sigma_2 + \sigma_3 + \\
+\theta\sigma_3)\Big) + 3\kappa_2\mu_2\Big(\mu_1\Big(- 3(1-2\theta)\sigma_1 + 2(3+2\theta)\sigma_2 - \\
-(3+4\theta)\sigma_3\Big) + \mu_2\Big((5-6\theta)\sigma_1 + 4(1-\theta)\sigma_2 - (3-4\theta)\sigma_3\Big)\Big)\Big) \geq 0,
\end{aligned}
\tag{3.63}
$$

and either

$$
\begin{aligned}
-6\theta^2\delta\kappa\delta\mu^2(3\kappa_2 + \mu_2)\sigma_1 - \theta\delta\mu\Big(\kappa_1\kappa_2\Big(9\mu_1(\sigma_3 - \sigma_1) + 33\mu_2\sigma_1 + 6\mu_2\sigma_2 - \\
-3\mu_2\sigma_3\Big) - 6\kappa_2^2\Big(6\mu_2\sigma_1 + \mu_1(-2\sigma_1 + \sigma_2 + \sigma_3)\Big) + 2\kappa_1\mu_2\Big(\mu_2(4\sigma_1 + \sigma_2 + \\
+\sigma_3) + 3\mu_1(\sigma_1 + 2\sigma_3)\Big) - 2\kappa_2\mu_2\Big(6\mu_2\sigma_1 + \mu_1(\sigma_1 + \sigma_2 + 7\sigma_3)\Big)\Big) + \\
+\mu_2\Big(\kappa_1\Big(3\kappa_2\mu_2(5\sigma_1 + 2\sigma_2 - \sigma_3) + 9\kappa_2\mu_1(-\sigma_1 + \sigma_3) + 2\mu_2(3\mu_1 + \\
+\mu_2)(\sigma_1 + \sigma_2 + \sigma_3)\Big) + 2\kappa_2\Big(- 3\mu_2^2\sigma_1 + 3\mu_1^2(-\sigma_2 + \sigma_3) - \mu_1\mu_2(\sigma_1 + \\
+\sigma_2 + 7\sigma_3) - 3\kappa_2\Big(3\mu_2\sigma_1 + \mu_1(-2\sigma_1 + \sigma_2 + \sigma_3)\Big)\Big)\Big)\Big) \geq 0,
\end{aligned}
\tag{3.64}
$$

or

$$\begin{aligned}
&-6\theta^2\delta\kappa\delta\mu^2(3\kappa_2+\mu_2)\sigma_3\theta\delta\mu\Big(-2\kappa_2\Big(3\kappa_2\Big(\mu_1(\sigma_1+\sigma_2-2\sigma_3)+6\mu_2\sigma_3\Big)+\\
&+\mu_2\Big(6\mu_2\sigma_3+\mu_1(7\sigma_1+\sigma_2+\sigma_3)\Big)\Big)+\kappa_1\Big(\kappa_2(9\mu_1\sigma_1-3\mu_2\sigma_1+6\mu_2\sigma_2-\\
&-9\mu_1\sigma_3+33\mu_2\sigma_3)+2\mu_2\Big(3\mu_1(2\sigma_1+\sigma_3)+\mu_2(\sigma_1+\sigma_2+4\sigma_3)\Big)\Big)\Big)+\\
&+\mu_2\Big(-2\kappa_2\Big(3\mu_1^2(-\sigma_1+\sigma_2)+3\mu_2^2\sigma_3+\mu_1\mu_2(7\sigma_1+\sigma_2+\sigma_3)+\\
&+3\kappa_2\Big(\mu_1(\sigma_1+\sigma_2-2\sigma_3)+3\mu_2\sigma_3\Big)\Big)+\kappa_1\Big(2\mu_2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\\
&+\sigma_3)+3\kappa_2\Big(3\mu_1(\sigma_1-\sigma_3)+\mu_2(-\sigma_1+2\sigma_2+5\sigma_3)\Big)\Big)\Big)\le 0,
\end{aligned} \tag{3.65}$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}\ge\mathbf{A}_2^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}+\frac{\theta}{\zeta_4}f_4(\sigma_1,\sigma_2,\sigma_3). \tag{3.66}$$

H. If

$$\begin{aligned}
&-6\theta^2\delta\kappa\delta\mu\mu_2\sigma_1+\mu_2\Big(-6\kappa_2\mu_2\sigma_1+3\kappa_1\kappa_2(-\sigma_1+\sigma_2)+2\kappa_2\mu_1(\sigma_2-\\
&-\sigma_3)+2\kappa_1\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)+\theta\Big(2\kappa_2\mu_2\Big(6\mu_2\sigma_1+\mu_1(-3\sigma_1-\sigma_2+\\
&+\sigma_3)\Big)+\kappa_1\Big(3\kappa_2\delta\mu(\sigma_1-\sigma_2)+2\mu_1\mu_2(\sigma_1+2\sigma_2)-\\
&-2\mu_2^2(4\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\ge 0,\\
&-6\theta^2\delta\kappa\delta\mu^2(3\kappa_2+\mu_2)\sigma_1-\theta\delta\mu\Big(\kappa_1\kappa_2\Big(9\mu_1(\sigma_3-\sigma_1)+33\mu_2\sigma_1+6\mu_2\sigma_2-\\
&-3\mu_2\sigma_3\Big)-6\kappa_2^2\Big(6\mu_2\sigma_1+\mu_1(-2\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_1\mu_2\Big(\mu_2(4\sigma_1+\\
&+\sigma_2+\sigma_3)+3\mu_1(\sigma_1+2\sigma_3)\Big)-2\kappa_2\mu_2\Big(6\mu_2\sigma_1+\mu_1(\sigma_1+\sigma_2+7\sigma_3)\Big)\Big)+\\
&+\mu_2\Big(\kappa_1\Big(3\kappa_2\mu_2(5\sigma_1+2\sigma_2-\sigma_3)+9\kappa_2\mu_1(-\sigma_1+\sigma_3)+2\mu_2(3\mu_1+\\
&+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\Big(-3\mu_2^2\sigma_1+3\mu_1^2(-\sigma_2+\sigma_3)-\mu_1\mu_2(\sigma_1+\\
&+\sigma_2+7\sigma_3)-3\kappa_2\Big(3\mu_2\sigma_1+\mu_1(-2\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\Big)\le 0,
\end{aligned} \tag{3.67}$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}\ge\mathbf{A}_2^{-1}\boldsymbol{\sigma}:\boldsymbol{\sigma}+\frac{4\theta}{\zeta_3(\theta\delta\mu-\mu_2)}f_3(\sigma_1,\sigma_2,\sigma_3). \tag{3.68}$$

I. If

$$\begin{aligned}
&\sigma_1\ge\sigma_2,\\
&12(1-\theta)\delta\kappa\mu_2\sigma_3-\kappa_1(3\kappa_2+4\mu_2)(\sigma_1+\sigma_2-2\sigma_3)\ge 0,\\
&(3\kappa_2+4\mu_2)\Big(-4\kappa_2\mu_1+\kappa_1(3\mu_1+\mu_2)\Big)(\sigma_1+\sigma_2)+2\Big(3\kappa_2\Big((-3+\\
&+6\theta)\kappa_1+(4-6\theta)\kappa_2\Big)\mu_1+\Big(3\kappa_2\Big((5-6\theta)\kappa_1+6(-1+\theta)\kappa_2\Big)+2\Big(3(1+\\
&+\theta)\kappa_1-\kappa_2-3\theta\kappa_2\Big)\mu_1\Big)\mu_2+2\Big(\kappa_1-3\theta\kappa_1+3(-1+\theta)\kappa_2\Big)\mu_2^2\Big)\sigma_3\ge 0,
\end{aligned} \tag{3.69}$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{\theta}{\zeta_3} f_2(\sigma_3, \sigma_2, \sigma_1). \tag{3.70}$$

J. If

$$\begin{aligned}
&\Big(\kappa_1(3\kappa_2 - 2\mu_2) + 6\mu_2(\kappa_2 - \theta\delta\kappa)\Big)(-\sigma_1 + \sigma_2 + \sigma_3) - \kappa_1(3\kappa_2 + 4\mu_2)\sigma_1 \geq 0,\\
&2\kappa_2\Big(- 9(1-\theta)\kappa_2\mu_2(\sigma_2+\sigma_3) + 3(-1+\theta)\mu_2^2(3\sigma_1 + \sigma_2 + \sigma_3) +\\
&+3\kappa_2\mu_1\Big(- 2\sigma_1 + (1-3\theta)(\sigma_2+\sigma_3)\Big) - \mu_1\mu_2\Big((-1+9\theta)\sigma_1 + (5+\\
&+3\theta)(\sigma_2+\sigma_3)\Big)\Big) + \kappa_1\Big(3\kappa_2\Big(6\mu_1\sigma_1 - 2\mu_2\sigma_1 + 3(-1+2\theta)\mu_1(\sigma_2+\sigma_3)+\\
&+(7-6\theta)\mu_2(\sigma_2+\sigma_3)\Big) + 2\mu_2\Big(3\mu_1\Big(\sigma_1 + 3\theta\sigma_1 + (1+\theta)(\sigma_2+\sigma_3)\Big)+\\
&+\mu_2\Big((5-9\theta)\sigma_1 + (5-3\theta)(\sigma_2+\sigma_3)\Big)\Big)\Big) \geq 0,\\
&2\kappa_1\mu_2\Big((1-\theta)\mu_2(\sigma_1+\sigma_2+\sigma_3-3\theta\sigma_3) + \theta\mu_1(2\sigma_2+\sigma_3-3\theta\sigma_3)\Big)+\\
&+\kappa_2\Big(3(-1+\theta)\mu_2\Big(2(1-\theta)\mu_2\sigma_3 + \kappa_1(-\sigma_2+\sigma_3)\Big)+\\
&+\mu_1\Big(3\theta\kappa_1(\sigma_2-\sigma_3) + 2(-1+\theta)\mu_2(\sigma_1-\sigma_2+3\theta\sigma_3)\Big)\Big) \geq 0,
\end{aligned} \tag{3.71}$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{\theta}{\zeta_2} f_1(\sigma_3, \sigma_2, \sigma_1). \tag{3.72}$$

K. If

$$\begin{aligned}
&2\kappa_1\mu_2\Big((1-\theta)\mu_2(\sigma_1+\sigma_2+\sigma_3-3\theta\sigma_3) + \theta\mu_1(2\sigma_2+\sigma_3-3\theta\sigma_3)\Big)+\\
&+\kappa_2\Big(- 3(1-\theta)\mu_2\Big(2(1-\theta)\mu_2\sigma_3 + \kappa_1(-\sigma_2+\sigma_3)\Big)+\\
&+\mu_1\Big(3\theta\kappa_1(\sigma_2-\sigma_3) + 2(-1+\theta)\mu_2(\sigma_1-\sigma_2+3\theta\sigma_3)\Big)\Big) \leq 0,\\
&-6\theta^2\delta\kappa\delta\mu^2(3\kappa_2+\mu_2)\sigma_3 + \theta(\mu_1-\mu_2)\Big(- 2\kappa_2\Big(3\kappa_2\Big(\mu_1(\sigma_1+\sigma_2-2\sigma_3)+\\
&+6\mu_2\sigma_3\Big) + \mu_2\Big(6\mu_2\sigma_3 + \mu_1(7\sigma_1+\sigma_2+\sigma_3)\Big)\Big) + \kappa_1\Big(\kappa_2(9\mu_1\sigma_1 - 3\mu_2\sigma_1+\\
&+6\mu_2\sigma_2 - 9\mu_1\sigma_3 + 33\mu_2\sigma_3) + 2\mu_2\Big(3\mu_1(2\sigma_1+\sigma_3) + \mu_2(\sigma_1+\sigma_2+\\
&+4\sigma_3)\Big)\Big)\Big) + \mu_2\Big(- 2\kappa_2\Big(3\mu_1^2(-\sigma_1+\sigma_2) + 3\mu_2^2\sigma_3 + \mu_1\mu_2(7\sigma_1+\sigma_2+\\
&+\sigma_3)3\kappa_2(\mu_1(\sigma_1+\sigma_2-2\sigma_3) + 3\mu_2\sigma_3)\Big) + \kappa_1\Big(2\mu_2(3\mu_1+\mu_2)(\sigma_1+\\
&+\sigma_2+\sigma_3) + 3\kappa_2\Big(3\mu_1(\sigma_1-\sigma_3) + \mu_2(-\sigma_1+2\sigma_2+5\sigma_3)\Big)\Big)\Big) \geq 0,
\end{aligned} \tag{3.73}$$

then

$$\mathbf{A}^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} \geq \mathbf{A}_2^{-1}\boldsymbol{\sigma} : \boldsymbol{\sigma} + \frac{4\theta}{(\theta\delta\mu - \mu_2)\zeta_3} f_3(\sigma_3, \sigma_2, \sigma_1). \tag{3.74}$$

Proof. In order to explicitly calculate the bound, we need to solve the following maximization problem

$$F(\boldsymbol{\eta}) = 2\boldsymbol{\sigma} : \boldsymbol{\eta} - (\mathbf{A}_1^{-1} - \mathbf{A}_2^{-1})^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} - (1-\theta)g^c(\boldsymbol{\eta}) \longrightarrow \max_{\boldsymbol{\eta}\in\mathrm{Sym}_3}.$$

Recalling Theorem 2.5, the above maximization is equvalent to maximization of the function

$$\begin{aligned} F(\eta_1,\eta_2,\eta_3) &= 2(\sigma_1\eta_1+\sigma_2\eta_2+\sigma_3\eta_3) - \frac{2\mu_1\mu_2}{\delta\mu}(\eta_1^2+\eta_2^2+\eta_3^2)- \\ &- \left(\frac{\kappa_1\kappa_2}{\delta\kappa} - \frac{2\mu_1\mu_2}{3\delta\mu}\right)(\eta_1+\eta_2+\eta_3)^2 - (1-\theta)g^c(\eta_1,\eta_2,\eta_3), \end{aligned} \tag{3.75}$$

over all pairs $(\eta_1,\eta_2,\eta_3) \in \mathbf{R}^3$. The function F is continuous and concave, but not necessary smooth, since g^c is not smooth. Notice that since $\sigma_1 \le \sigma_2 \le \sigma_3$, it follows that for a maximizer $(\eta_1^*,\eta_2^*,\eta_3^*)$ of the function F the inequalities $\eta_1^* \le \eta_2^* \le \eta_3^*$ are valid. If not, then for $\eta_2^* < \eta_1^* \le \eta_3^*$, it follows that $F(\eta_1^*,\eta_2^*,\eta_3^*) < F(\eta_2^*,\eta_1^*,\eta_3^*)$ (analogous inequality follows if some other inequality in $\eta_1^* \le \eta_2^* \le \eta_3^*$ is not satisfied). Therefore, we have to solve the following constrained maximization problem

$$\begin{cases} F(\eta_1,\eta_2,\eta_3) = 2(\sigma_1\eta_1+\sigma_2\eta_2+\sigma_3\eta_3) - \dfrac{2\mu_1\mu_2}{\delta\mu}(\eta_1^2+\eta_2^2+\eta_3^2)- \\ -\left(\dfrac{\kappa_1\kappa_2}{\delta\kappa} - \dfrac{2\mu_1\mu_2}{3\delta\mu}\right)(\eta_1+\eta_2+\eta_3)^2 - (1-\theta)\Big(2\mu_2(\eta_1^2+\eta_2^2+\eta_3^2)+\lambda_2(\eta_1+\eta_2+\eta_3)^2- \\ -\dfrac{1}{2\mu_2+\lambda_2}\min\{(2\mu_2\eta_1+\lambda_2\mathrm{tr}\boldsymbol{\eta})^2,(2\mu_2\eta_3+\lambda_2\mathrm{tr}\boldsymbol{\eta})^2\}\Big) \longrightarrow \max \\ (\eta_1,\eta_2,\eta_3) \in \mathbf{R}^3, \\ \eta_1 \le \eta_2 \le \eta_3. \end{cases}$$

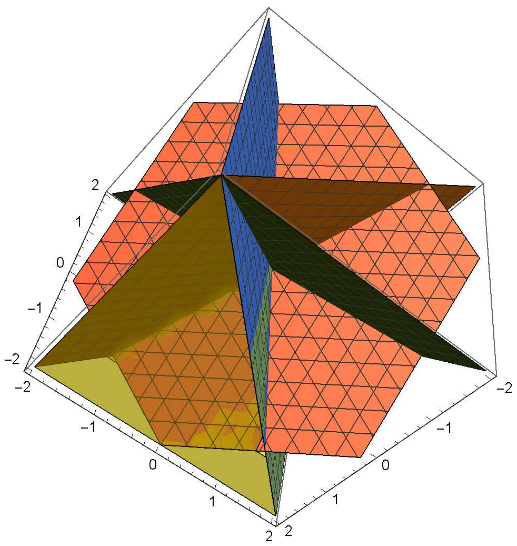


Figure 3.7: A region in $\mathbf{R}^3$ over which the function F is maximized (shaded with the yellow tones), for $\mu_2 = 1$, $\lambda_2 = 4$. The green plane is plane $\eta_1 = \eta_3$, the blue one is $\eta_1 = \eta_2$, the red one is $\eta_2 = \eta_3$, while the orange one is $\mu_2(\eta_1+\eta_3)+\lambda_2\mathrm{tr}\boldsymbol{\eta} = 0$.

In Figure 3.7 a set by which we maximize the function F is given. The orange plane represents a plane $\mu_2(\eta_1+\eta_3)+\lambda_2\mathrm{tr}\boldsymbol{\eta} = 0$, which comes from the equality $(2\mu_2\eta_1+\lambda_2\mathrm{tr}\boldsymbol{\eta})^2 =$

$(2\mu_2\eta_3 + \lambda_2 \mathrm{tr}\boldsymbol{\eta})^2$. The function F is a piecewise function, defined by different terms on every side of the plane. Therefore, we distinguish two cases: $\mu_2(\eta_1+\eta_3)+\lambda_2\mathrm{tr}\boldsymbol{\eta} \geq 0$, and $\mu_2(\eta_1+\eta_3)+\lambda_2\mathrm{tr}\boldsymbol{\eta} \leq 0$.

I. Let us assume that $\mu_2(\eta_1+\eta_3)+\lambda_2\mathrm{tr}\boldsymbol{\eta} \geq 0$. In this case, $\min\{(2\mu_2\eta_1+\lambda_2\mathrm{tr}\boldsymbol{\eta})^2, (2\mu_2\eta_3+\lambda_2\mathrm{tr}\boldsymbol{\eta})^2\} = (2\mu_2\eta_1+\lambda_2\mathrm{tr}\boldsymbol{\eta})^2$, and the maximization problem reads

$$\begin{cases} F(\eta_1,\eta_2,\eta_3) = 2(\sigma_1\eta_1+\sigma_2\eta_2+\sigma_3\eta_3) - \dfrac{2\mu_1\mu_2}{\delta\mu}(\eta_1^2+\eta_2^2+\eta_3^2) - \\ -\left(\dfrac{\kappa_1\kappa_2}{\delta\kappa} - \dfrac{2\mu_1\mu_2}{3\delta\mu}\right)(\eta_1+\eta_2+\eta_3)^2 - (1-\theta)\Big(2\mu_2(\eta_1^2+\eta_2^2+\eta_3^2) + \lambda_2(\eta_1+\eta_2+\eta_3)^2 - \\ -\dfrac{1}{2\mu_2+\lambda_2}(2\mu_2\eta_1+\lambda_2(\eta_1+\eta_2+\eta_3))^2\Big) \longrightarrow \max, \\ \eta_1-\eta_2 \leq 0, \\ \eta_2-\eta_3 \leq 0, \\ -(\mu_2+\lambda_2)(\eta_1+\eta_3) - \lambda_2\eta_2 \leq 0. \end{cases}$$

We shall solve the above problem by the Karush-Kuhn-Tucker (KKT) system, which reads

$$\begin{cases} -2\sigma_1 + \dfrac{4\mu_1\mu_2}{\delta\mu}\eta_1 + 2\alpha(\eta_1+\eta_2+\eta_3) + a_1 - a_3(\mu_2+\lambda_2) = 0, \\ -2\sigma_2 + \dfrac{4\mu_1\mu_2}{\delta\mu}\eta_2 + 2\alpha(\eta_1+\eta_2+\eta_3) + (1-\theta)(4\mu_2\eta_2 + \beta(\eta_2+\eta_3)) - a_1 + a_2 - a_3\lambda_2 = 0, \\ -2\sigma_3 + \dfrac{4\mu_1\mu_2}{\delta\mu}\eta_3 + 2\alpha(\eta_1+\eta_2+\eta_3) + (1-\theta)(4\mu_2\eta_3 + \beta(\eta_2+\eta_3)) - a_2 - a_3(\mu_2+\lambda_2) = 0, \\ a_1(\eta_1-\eta_2) = 0, \\ a_2(\eta_2-\eta_3) = 0, \\ a_3((\mu_2+\lambda_2)(\eta_1+\eta_3)+\lambda_2\eta_2) = 0, \\ \eta_1-\eta_2 \leq 0, \\ \eta_2-\eta_3 \leq 0, \\ -(\mu_2+\lambda_2)(\eta_1+\eta_3) - \lambda_2\eta_2 \leq 0, \\ a_1, a_2, a_3 \geq 0, \end{cases}$$

where $\alpha = \frac{\kappa_1\kappa_2}{\delta\kappa} - \frac{2\mu_1\mu_2}{3\delta\mu}$, and $\beta = \frac{4\mu_2\lambda_2}{2\mu_2+\lambda_2}$. There are three complementary conditions, and therefore, we need to check 8 cases.

Case 1. If $\eta_1 = \eta_2 = \eta_3$ and $(\mu_2+\lambda_2)(\eta_1+\eta_3)+\lambda_2\eta_2 = 0$, then $\eta_1=\eta_2=\eta_3=0$, which implies that $F(\eta_1,\eta_2,\eta_3) = 0$. We calculate parameters

$$a_1 = \frac{6\kappa_2(2\sigma_1-\sigma_2-\sigma_3) - 2\mu_2(\sigma_1+\sigma_2+\sigma_3)}{9\kappa_2},$$

$$a_2 = \frac{2\Big(3\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)}{9\kappa_2},\ a_3 = -\frac{2(\sigma_1+\sigma_2+\sigma_3)}{3\kappa_2}$$

and, since $\kappa_2 > 0$, we obtain conditions

$$\begin{aligned} \sigma_1+\sigma_2+\sigma_3 &\le 0 \\ 3\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3) &\ge 0 \\ 3\kappa_2(2\sigma_1-\sigma_2-\sigma_3)-\mu_2(\sigma_1+\sigma_2+\sigma_3) &\ge 0. \end{aligned}$$

The last two conditions are (3.46), but here we have an additional constraint on the sum of eigenvalues of the matrix $\boldsymbol{\sigma}$.

Case 2. If $\eta_1 = \eta_2 = \eta_3$ and $a_3 = 0$, then we calculate

$$a_1 = -\frac{2\Big(\kappa_1\Big(4\mu_2(-3\theta\sigma_1+\sigma_1+\sigma_2+\sigma_3)+3\kappa_2(-2\sigma_1+\sigma_2+\sigma_3)\Big)-12(1-\theta)\kappa_2\mu_2\sigma_1\Big)}{3\Big(4(1-\theta)\kappa_2\mu_2+\kappa_1(3\kappa_2+4\theta\mu_2)\Big)},$$

$$\begin{aligned} a_2 =& \frac{2\kappa_1(\sigma_1+\sigma_2)(6\theta\mu_2+3\kappa_2-2\mu_2)-4\kappa_1\sigma_3(3\theta\mu_2+3\kappa_2+\mu_2)}{3\Big(4(1-\theta)\kappa_2\mu_2+\kappa_1(3\kappa_2+4\theta\mu_2)\Big)}+ \\ &+\frac{4(1-\theta)\kappa_2\mu_2(\sigma_1+\sigma_2-\sigma_3)}{4(1-\theta)\kappa_2\mu_2+\kappa_1(3\kappa_2+4\theta\mu_2)}, \end{aligned}$$

and

$$\eta_1=\eta_2=\eta_3 = \frac{\delta\kappa(3\kappa_2+4\mu_2)(\sigma_1+\sigma_2+\sigma_3)}{9\kappa_2\Big(4(1-\theta)\kappa_2\mu_2+\kappa_1(3\kappa_2+4\theta\mu_2)\Big)}.$$

The maximum value in this case is

$$F(\eta_1,\eta_2,\eta_3) = \frac{\delta\kappa(3\kappa_2+4\mu_2)(\sigma_1+\sigma_2+\sigma_3)^2}{9\kappa_2\Big(4(1-\theta)\kappa_2\mu_2+\kappa_1(3\kappa_2+4\theta\mu_2)\Big)},$$

and threfore, we obtain the bound (3.50). The condition $-(\mu_2+\lambda_2)(\eta_1+\eta_3)-\lambda_2\eta_2 \le 0$ is equivalent to

$$\sigma_1+\sigma_2+\sigma_3 \ge 0,$$

which is the first condition in (3.48), while the other two follow from nonnegativity of a_1 and a_2, since $4(1-\theta)\kappa_2\mu_2+\kappa_1(3\kappa_2+4\theta\mu_2) \ge 0$.

Case 3. If $\eta_1 = \eta_2$, $(\mu_2+\lambda_2)(\eta_1+\eta_3)+\lambda_2\eta_2 = 0$ and $a_2 = 0$, then

$$\begin{aligned} a_1 = \frac{18}{\zeta_1}\Big(&2\theta\delta\kappa\delta\mu(3\kappa_2+\mu_2)\Big((3\kappa_2+\mu_2)(\sigma_1-\sigma_2)-3\mu_2\sigma_3\Big)-\mu_2\Big(\kappa_1\Big(18\kappa_2^2(\sigma_2-\sigma_1)+ \\ &+3\kappa_2\Big(\mu_2(-3\sigma_1+4\sigma_2+5\sigma_3)-3\mu_1(\sigma_1-2\sigma_2+\sigma_3)\Big)+2\mu_2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)+ \\ &+2\kappa_2\Big(9\kappa_2^2(\sigma_1-\sigma_2)+3\kappa_2\Big(\mu_1(\sigma_1-3\sigma_2+2\sigma_3)+\mu_2(2\sigma_1-2\sigma_2-3\sigma_3)\Big)+ \\ &+\mu_2\Big(\mu_2(\sigma_1-\sigma_2-3\sigma_3)-\mu_1(5\sigma_1+3\sigma_2+\sigma_3)\Big)\Big)\Big)\Big), \end{aligned}$$

$$
\begin{aligned}
a_3 = \frac{-18}{\zeta_1}\Big(&\kappa_1(\sigma_1+\sigma_2)\Big(3\kappa_2\Big((6\theta-3)\mu_1+(7-6\theta)\mu_2\Big)+2\mu_2\Big(3(\theta+1)\mu_1+\\
&+(5-3\theta)\mu_2\Big)\Big)+2\kappa_1\sigma_3\Big(\mu_2\Big((9\theta+3)\mu_1+(5-9\theta)\mu_2\Big)+\kappa_2(9\mu_1-3\mu_2)\Big)-\\
&-2\kappa_2\Big(-3\kappa_2(\sigma_1+\sigma_2)\Big(-3\theta\mu_1+3(\theta-1)\mu_2+\mu_1\Big)+\mu_2\Big((3\theta+5)\mu_1(\sigma_1+\sigma_2)+\\
&+(9\theta-1)\mu_1\sigma_3-3(\theta-1)\mu_2(\sigma_1+\sigma_2+3\sigma_3)\Big)+6\kappa_2\mu_1\sigma_3\Big)\Big),
\end{aligned}
$$

and

$$
\eta_1 = \frac{-1}{\zeta_1}\delta\kappa\delta\mu(3\kappa_2+\mu_2)(3\kappa_2+4\mu_2)\Big(3\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big).
$$

By introducing η_1, $\eta_2=\eta_1$ and $\eta_3=-\dfrac{\mu_2+2\lambda_2}{\mu_2+\lambda_2}\eta_1$ into the function F, we obtain

$$
F(\eta_1,\eta_2,\eta_3) = \frac{-1}{\zeta_1}\delta\kappa\delta\mu(3\kappa_2+4\mu_2)\Big(3\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)^2,
$$

which gives the bound (3.54). The condition $\eta_2-\eta_3\le 0$ is equivalent to

$$
3\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3)\le 0,
$$

which, together with condition $a_1\ge 0$ and $a_3\ge 0$ gives (3.51) and (3.52), since $\zeta_1\le 0$.

Case 4. If $\eta_1=\eta_2$ and $a_2=a_3=0$, then

$$
\eta_1=\frac{\xi_1}{\zeta_2},\qquad \eta_3=\frac{2\xi_3}{\zeta_2},
$$

where

$$
\begin{aligned}
\xi_1 =\ & 2\kappa_2\mu_2\Big(6\kappa_2(\sigma_1+\sigma_2)\Big((3\theta-2)\mu_1+3(1-\theta)\mu_2\Big)+3\kappa_2\sigma_3\Big((5-3\theta)\mu_1-3(1-\theta)\mu_2\Big)+\\
&+2\mu_2\Big(3\theta\mu_1-3(\theta-1)\mu_2+\mu_1\Big)(\sigma_1+\sigma_2+\sigma_3)\Big)-\kappa_1\Big(6\kappa_2\mu_2\Big(\sigma_3(-3\theta\mu_1+3\theta\mu_2+\mu_1+\mu_2)-\\
&-2(\sigma_1+\sigma_2)(-3\theta\mu_1+3\theta\mu_2+\mu_1-2\mu_2)\Big)+4\mu_2^2(-3\theta\delta\mu+\mu_1+3\mu_2)(\sigma_1+\sigma_2+\sigma_3)-\\
&-9\kappa_2^2\delta\mu(\sigma_1+\sigma_2-2\sigma_3)\Big),
\end{aligned}
$$

and

$$
\begin{aligned}
\xi_3 = \kappa_1\Big(&-3\kappa_2\mu_2(\sigma_1+\sigma_2)(3\theta\delta\mu+\mu_1+\mu_2)+6\kappa_2\mu_2\sigma_3\Big(-(3\theta+2)\mu_1+3\theta\mu_2+\mu_2\Big)-\\
&-2\mu_2^2(3\theta\mu_1+3(1-\theta)\mu_2+\mu_1)(\sigma_1+\sigma_2+\sigma_3)-9\kappa_2^2\delta\mu(\sigma_1+\sigma_2-2\sigma_3)\Big)+\\
&+\kappa_2\mu_2\Big(-3(3\theta-5)\kappa_2\mu_1(\sigma_1+\sigma_2)+6(3\theta-2)\kappa_2\mu_1\sigma_3+9(\theta-1)\kappa_2\mu_2(\sigma_1+\sigma_2-2\sigma_3)+\\
&+2\mu_2\Big(3\theta\mu_1+3(1-\theta)\mu_2+\mu_1\Big)(\sigma_1+\sigma_2+\sigma_3)\Big),
\end{aligned}
$$

while

$$a_1 = \frac{36}{\zeta_2}\Big(6\theta^2\mu_2\delta\kappa\delta\mu\sigma_1 + \theta\Big(\kappa_1\Big(3\kappa_2\delta\mu(\sigma_2-\sigma_1) + 2\mu_2\Big(\mu_2(4\sigma_1+\sigma_2+\sigma_3)-$$
$$-\mu_1(\sigma_1+2\sigma_2)\Big)\Big) + 2\kappa_2\mu_2\Big(\mu_1(3\sigma_1+\sigma_2-\sigma_3) - 6\mu_2\sigma_1\Big)\Big) - \mu_2\Big(\kappa_1\Big(3\kappa_2(\sigma_2-\sigma_1)+$$
$$+2\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big) - 2\kappa_2\Big(\mu_1(\sigma_3-\sigma_2) + 3\mu_2\sigma_1\Big)\Big)\Big).$$

The maximum in this case is

$$F(\eta_1,\eta_2,\eta_3) = \frac{f_1(\sigma_1,\sigma_2,\sigma_3)}{\zeta_2},$$

which gives the Hashin-Shtrikman bound (3.56) in case D of the theorem. Inequalities from the condition (3.55) are easily obtained from conditions $\eta_2 - \eta_3 \le 0$, which is equivalent to

$$\Big(\kappa_1(3\kappa_2 - 2\mu_2) + 6\mu_2(\kappa_2 - \theta\delta\kappa)\Big)(\sigma_1+\sigma_2-\sigma_3) - \kappa_1(3\kappa_2+4\mu_2)\sigma_3 \le 0,$$

and $-\mu_2(\eta_1+\eta_3) - \lambda_2(\eta_1+\eta_2+\eta_3) \le 0$, which is equivalent to

$$-2\kappa_2\Big((-3+9\theta)\kappa_2\mu_1 - 9(-1+\theta)\kappa_2\mu_2 + (5+3\theta)\mu_1\mu_2 - 3(-1+\theta)\mu_2^2\Big)(\sigma_1+\sigma_2)+$$
$$+\kappa_1\Big(3\kappa_2\Big((-3+6\theta)\mu_1 + (7-6\theta)\mu_2\Big) + 2\mu_2\Big(3(1+\theta)\mu_1 + (5-3\theta)\mu_2\Big)\Big)(\sigma_1+\sigma_2)+$$
$$+2\kappa_1\Big(\kappa_2(9\mu_1 - 3\mu_2) + \mu_2\Big((3+9\theta)\mu_1 + (5-9\theta)\mu_2\Big)\Big)\sigma_3-$$
$$-2\kappa_2\Big(6\kappa_2\mu_1 - \mu_2\Big(\mu_1 - 9\theta\mu_1 + 9(-1+\theta)\mu_2\Big)\Big)\sigma_3 < 0,$$

and from $a_1 \ge 0$, since $\zeta_2 \ge 0$.

Case 5. Let $a_1 = 0$, $\eta_2 = \eta_3$ and $(\mu_2+\lambda_2)(\eta_1+\eta_3) + \lambda_2\eta_2 = 0$. Then

$$\eta_2 = \frac{1}{\zeta_1}\delta\kappa\delta\mu(3\kappa_2+\mu_2)(3\kappa_2+4\mu_2)\Big(\kappa_2\Big(6\sigma_1 - 3(\sigma_2+\sigma_3)\Big) - \mu_2(\sigma_1+\sigma_2+\sigma_3)\Big),$$

$$a_2 = \frac{18\kappa_2\mu_2}{\zeta_1}\Big(\kappa_1\Big(18\kappa_2^2(\sigma_2-\sigma_3)(-\theta\delta\mu+\mu_2) + 3\kappa_2\mu_2\Big(\mu_1(6\theta\sigma_1 + 4\theta\sigma_2 - 4\theta\sigma_3 - 3\sigma_1+$$
$$+6\sigma_2 - 3\sigma_3) + \mu_2(-6\theta\sigma_1 - 4\theta\sigma_2 + 4\theta\sigma_3 + 5\sigma_1 + 4\sigma_2 - 3\sigma_3)\Big) + 2\mu_2^2\Big(\mu_1(3(\theta+$$
$$+1)\sigma_1 + (\theta+3)\sigma_2 - (\theta-3)\sigma_3) + \mu_2(-3\theta\sigma_1 - \theta\sigma_2 + \theta\sigma_3 + \sigma_1 + \sigma_2 + \sigma_3)\Big)\Big)-$$
$$-2\kappa_2\Big(9\kappa_2^2(\sigma_2-\sigma_3)(-\theta\delta\mu+\mu_2) - 3\kappa_2\mu_2\Big(\mu_1\Big((2-3\theta)\sigma_1 - (3+2\theta)\sigma_2+$$
$$+(1+2\theta)\sigma_3\Big) + (\theta-1)\mu_2(3\sigma_1 + 2\sigma_2 - 2\sigma_3)\Big) + \mu_2^2\Big(\mu_1\Big(3\theta\sigma_1 + (\theta+3)\sigma_2-$$
$$-(\theta-5)\sigma_3 + \sigma_1\Big) + (\theta-1)\mu_2(-3\sigma_1 - \sigma_2 + \sigma_3)\Big)\Big)\Big),$$

and

$$\begin{aligned}a_3 = \frac{-18\kappa_2\mu_2}{\zeta_1}\Big(\kappa_1\Big(3\kappa_2\Big(3\mu_1(4\theta\sigma_1 - 2\sigma_1 + \sigma_2 + \sigma_3) + \mu_2\Big(2(5-6\theta)\sigma_1 + \sigma_2 + \sigma_3\Big)\Big) +\\ + 4\mu_2\Big(3\mu_1(\theta\sigma_1 + \sigma_1 + \sigma_2 + \sigma_3) + \mu_2(-3\theta\sigma_1 + \sigma_1 + \sigma_2 + \sigma_3)\Big)\Big) - 4\kappa_2\Big(3\kappa_2\Big(\mu_1(3\theta\sigma_1 -\\ - 2\sigma_1 + \sigma_2 + \sigma_3) + 3(1-\theta)\mu_2\sigma_1\Big) + \mu_2\Big(\mu_1(3\theta\sigma_1 + \sigma_1 + 4(\sigma_2+\sigma_3)\Big) + 3(1-\theta)\mu_2\sigma_1\Big)\Big)\Big).\end{aligned}$$

By introducing $\eta_1 = \dfrac{\mu_2 + 2\lambda_2}{\mu_2 + \lambda_2}\eta_2$, η_2 and $\eta_3 = \eta_2$ into the function F, we get

$$F(\eta_1, \eta_2, \eta_3) = \frac{-1}{\zeta_1}\delta\kappa\delta\mu(3\kappa_2 + 4\mu_2)\Big(3\kappa_2(-2\sigma_1 + \sigma_2 + \sigma_3) + \mu_2(\sigma_1 + \sigma_2 + \sigma_3)\Big)^2.$$

Condition $\eta_1 - \eta_2 \le 0$ is equivalent to

$$3\kappa_2(2\sigma_1 - \sigma_2 - \sigma_3) - \mu_2(\sigma_1 + \sigma_2 + \sigma_3) \le 0,$$

which together with nonnegativity conditions for parameters a_2 and a_3 gives conditions (3.57) and (3.58). The value of the function F in this case gives the bound (3.60) in case E of the theorem.

Case 6. If $\eta_2 = \eta_3$ and $a_1 = a_3 = 0$, then

$$\eta_1 = \frac{-2\xi_1}{\zeta_3},$$

where

$$\begin{aligned}\xi_1 =& \kappa_1\Big(-6\kappa_2\mu_2\Big(\mu_1\Big((4-9\theta)\sigma_1 + \sigma_2 + \sigma_3\Big) + (9\theta - 5)\mu_2\sigma_1 - 2\mu_2(\sigma_2 + \sigma_3)\Big) -\\ &-9\kappa_2^2\delta\mu(2\sigma_1 - \sigma_2 - \sigma_3) + 8\mu_1\mu_2^2(\sigma_1 + \sigma_2 + \sigma_3)\Big) - 2\kappa_2\mu_2\Big(3\kappa_2\Big(\mu_1(9\theta\sigma_1 - 8\sigma_1 +\\ &+\sigma_2 + \sigma_3) + 9(1-\theta)\mu_2\sigma_1\Big) + 4\mu_1\mu_2(\sigma_1 + \sigma_2 + \sigma_3)\Big),\end{aligned}$$

$$\eta_3 = \frac{(3\kappa_2 + 4\mu_2)}{\zeta_3}\Big(3\kappa_1\kappa_2\delta\mu(2\sigma_1 - \sigma_2 - \sigma_3) - 4\mu_1\mu_2\delta\kappa(\sigma_1 + \sigma_2 + \sigma_3)\Big),$$

and $a_2 = \sigma_2 - \sigma_3$. In this case, the maximum value reads

$$F = \frac{f_2(\sigma_1, \sigma_2, \sigma_3)}{\zeta_3}.$$

Condition $a_2 \ge 0$ is an assumption of the theorem, while condition $\eta_1 - \eta_2 \le 0$ is equivalent to

$$-12(1-\theta)\kappa_2\mu_2\sigma_1 + \kappa_1\Big(3\kappa_2(-2\sigma_1 + \sigma_2 + \sigma_3) + 4\mu_2(\sigma_1 - 3\theta\sigma_1 + \sigma_2 + \sigma_3)\Big) \ge 0$$

and condition $-\mu_2(\eta_1+\eta_3)-\lambda_2(\eta_1+\eta_2+\eta_3)\le 0$ is equivalent to

$$\begin{aligned}&-4\kappa_2\Big(-9(-1+\theta)\kappa_2\mu_2\sigma_1-3(-1+\theta)\mu_2^2\sigma_1+3\kappa_2\mu_1\Big((-2+3\theta)\sigma_1+\sigma_2+\sigma_3\Big)+\\&+\mu_1\mu_2\Big(\sigma_1+3\theta\sigma_1+4(\sigma_2+\sigma_3)\Big)\Big)+\kappa_1\Big(4\mu_2\Big(\mu_2(\sigma_1-3\theta\sigma_1+\sigma_2+\sigma_3)+3\mu_1(\sigma_1+\theta\sigma_1+\\&+\sigma_2+\sigma_3)\Big)+3\kappa_2\Big(\mu_2\Big(2(5-6\theta)\sigma_1+\sigma_2+\sigma_3\Big)+3\mu_1(-2\sigma_1+4\theta\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\le 0,\end{aligned}$$

which gives case F of the theorem.

Case 7. If $a_1=a_2=0$ and $(\mu_2+\lambda_2)(\eta_1+\eta_3)+\lambda_2\eta_2=0$, then

$$\eta_1=\frac{\xi_1}{\zeta_4},\quad \eta_3=\frac{\xi_3}{\zeta_4},$$

where

$$\begin{aligned}\xi_1=&-\delta\mu\Big(2\theta\delta\kappa\delta\mu(3\kappa_2+\mu_2)\Big(9\kappa_2^2(2\sigma_1-\sigma_2-\sigma_3)+3\kappa_2\mu_2(4\sigma_1+\sigma_2-8\sigma_3)+\\&+2\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)+\mu_2\Big(\kappa_1\Big(54\kappa_2^3(2\sigma_1-\sigma_2-\sigma_3)+27\kappa_2^2\Big(\mu_1(\sigma_1-2\sigma_2+\sigma_3)+\\&+\mu_2(3\sigma_1-5\sigma_3)\Big)+18\kappa_2\mu_2\Big(\mu_1(-2\sigma_1+\sigma_2-2\sigma_3)+\mu_2\sigma_2\Big)+4\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\\&+\sigma_3)\Big)-2\kappa_2\Big(27\kappa_2^3(2\sigma_1-\sigma_2-\sigma_3)+27\kappa_2^2(-\mu_1\sigma_2+\mu_1\sigma_3+2\mu_2\sigma_1-3\mu_2\sigma_3)+\\&+9\kappa_2\mu_2\Big(\mu_1(\sigma_2-4\sigma_1)+\mu_2(2\sigma_1+\sigma_2-2\sigma_3)\Big)+2\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\Big),\end{aligned}$$

and

$$\begin{aligned}\xi_3=&-\delta\mu\Big(\mu_2\Big(\kappa_1\Big(-54\kappa_2^3(\sigma_1+\sigma_2-2\sigma_3)+27\kappa_2^2(\mu_1(\sigma_1-2\sigma_2+\sigma_3)-5\mu_2\sigma_1+3\mu_2\sigma_3)+\\&+18\kappa_2\mu_2(\mu_1(-2\sigma_1+\sigma_2-2\sigma_3)+\mu_2\sigma_2)+4\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)+\\&+2\kappa_2\Big(27\kappa_2^3(\sigma_1+\sigma_2-2\sigma_3)+27\kappa_2^2(-\mu_1\sigma_1+\mu_1\sigma_2+3\mu_2\sigma_1-2\mu_2\sigma_3)-9\kappa_2\mu_2\Big(\mu_1(\sigma_2-\\&-4\sigma_3)+\mu_2(-2\sigma_1+\sigma_2+2\sigma_3)\Big)-2\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)\Big)-2\theta(3\kappa_2+\\&+\mu_2)\delta\mu\delta\kappa\Big(9\kappa_2^2(\sigma_1+\sigma_2-2\sigma_3)+3\kappa_2\mu_2(8\sigma_1-\sigma_2-4\sigma_3)-2\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)\Big),\end{aligned}$$

while

$$\begin{aligned}a_3=&\frac{36\kappa_2\mu_3}{\zeta_4}\Big(-6\theta^2\sigma_1\delta\kappa(3\kappa_2+\mu_2)\delta\mu^2-\theta\delta\mu\Big(\kappa_1\Big(\kappa_2\Big(9\mu_1(\sigma_3-\sigma_1)+33\mu_2\sigma_1+6\mu_2\sigma_2-\\&-3\mu_2\sigma_3\Big)+2\mu_2\Big(3\mu_1(\sigma_1+2\sigma_3)+\mu_2(4\sigma_1+\sigma_2+\sigma_3)\Big)\Big)+2\kappa_2\Big(-3\kappa_2\Big(\mu_1(-2\sigma_1+\sigma_2+\\&+\sigma_3)+6\mu_2\sigma_1\Big)-\mu_2\Big(\mu_1(\sigma_1+\sigma_2+7\sigma_3)+6\mu_2\sigma_1\Big)\Big)\Big)+\mu_2\Big(\kappa_1\Big(3\kappa_2\Big(3\mu_1(\sigma_3-\sigma_1)+\\&+\mu_2(5\sigma_1+2\sigma_2-\sigma_3)\Big)+2\mu_2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\Big(-3\kappa_2\Big(\mu_1(-2\sigma_1+\\&+\sigma_2+\sigma_3)+3\mu_2\sigma_1\Big)+3\mu_1^2(\sigma_3-\sigma_2)-\mu_1\mu_2(\sigma_1+\sigma_2+7\sigma_3)-3\mu_2^2\sigma_1\Big)\Big)\Big).\end{aligned}$$

By introducing η_1, $\eta_2 = -\dfrac{\mu_2+\lambda_2}{\lambda_2}(\eta_1+\eta_3)$ and η_3 into the function F, we obtain

$$F(\eta_1,\eta_2,\eta_3) = \frac{1}{\zeta_4} f_4(\sigma_1,\sigma_2,\sigma_3),$$

which gives the bound (3.66). Condition $\eta_1 - \eta_2 \le 0$ is equivalent to

$$\begin{aligned}
&-2\theta\delta\kappa\delta\mu(3\kappa_2+\mu_2)((3\kappa_2+\mu_2)(\sigma_1-\sigma_2)-3\mu_2\sigma_3)+\mu_2(2\kappa_2(9\kappa_2^2(\sigma_1-\sigma_2)+\\
&+\mu_2(\mu_2(\sigma_1-\sigma_2-3\sigma_3)-\mu_1(5\sigma_1+3\sigma_2+\sigma_3))+3\kappa_2(\mu_2(2\sigma_1-2\sigma_2-3\sigma_3)+\\
&+\mu_1(\sigma_1-3\sigma_2+2\sigma_3)))+\kappa_1(18\kappa_2^2(-\sigma_1+\sigma_2)+2\mu_2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)+\\
&+3\kappa_2(-3\mu_1(\sigma_1-2\sigma_2+\sigma_3)+\mu_2(-3\sigma_1+4\sigma_2+5\sigma_3)))) \le 0
\end{aligned}$$

and condition $\eta_2 - \eta_3 \le 0$ is equivalent to

$$\begin{aligned}
&-2\kappa_2\Big(9\kappa_2^2(-\theta\delta\mu+\mu_2)(\sigma_2-\sigma_3)+\mu_2^2\Big((1-\theta)\mu_2(3\sigma_1+\sigma_2-\sigma_3)+\mu_1\Big(\sigma_1+3\theta\sigma_1+\\
&+(3+\theta)\sigma_2+(5-\theta)\sigma_3\Big)\Big)-3\kappa_2\mu_2\Big((-1+\theta)\mu_2(3\sigma_1+2\sigma_2-2\sigma_3)+\mu_1(2\sigma_1-3\theta\sigma_1-\\
&-3\sigma_2-2\theta\sigma_2+\sigma_3+2\theta\sigma_3)\Big)\Big)+\kappa_1\Big(18\kappa_2^2(-\theta\delta\mu+\mu_2)(\sigma_2-\sigma_3)+2\mu_2^2\Big(\mu_1\Big(3(1+\theta)\sigma_1+\\
&+(3+\theta)\sigma_2+(3-\theta)\sigma_3\Big)+\mu_2(\sigma_1-3\theta\sigma_1+\sigma_2-\theta\sigma_2+\sigma_3+\theta\sigma_3)\Big)+3\kappa_2\mu_2\Big(\mu_1(-3\sigma_1+\\
&+6\theta\sigma_1+6\sigma_2+4\theta\sigma_2-3\sigma_3-4\theta\sigma_3)+\mu_2(5\sigma_1-6\theta\sigma_1+4\sigma_2-4\theta\sigma_2-3\sigma_3+4\theta\sigma_3)\Big)\Big) \ge 0.
\end{aligned}$$

These conditions give (3.63), while (3.64) is obtained from the conditon $a_3 \ge 0$.

Case 8. If $a_1 = a_2 = a_3 = 0$, then

$$\eta_1 = \frac{-2\xi_1}{\zeta_3}, \quad \eta_2 = \frac{2\xi_2}{(\theta\delta\mu-\mu_2)\zeta_3}, \quad \eta_3 = \frac{2\xi_3}{(\theta\delta\mu-\mu_2)\zeta_3},$$

where

$$\begin{aligned}
\xi_1 =&\kappa_1\Big(-6\kappa_2\mu_2\Big(\mu_1\Big((4-9\theta)\sigma_1+\sigma_2+\sigma_3\Big)+(9\theta-5)\mu_2\sigma_1-2\mu_2(\sigma_2+\sigma_3)\Big)-\\
&-9\kappa_2^2\delta\mu(2\sigma_1-\sigma_2-\sigma_3)+8\mu_1\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\mu_2\Big(-3\kappa_2\mu_1\Big((9\theta-8)\sigma_1+\\
&+\sigma_2+\sigma_3\Big)+27(\theta-1)\kappa_2\mu_2\sigma_1-4\mu_1\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big),
\end{aligned}$$

$$\begin{aligned}
\xi_2 =&-\theta\delta\mu\Big(\kappa_1\Big(-9\kappa_2^2\delta\mu(\sigma_1-2\sigma_2+\sigma_3)+6\kappa_2\mu_2\Big(\mu_1(\sigma_1-5\sigma_2+\sigma_3)+\mu_2(-2\sigma_1+\\
&+\sigma_2+\sigma_3)\Big)-8\mu_1\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\mu_1\mu_2\Big(3\kappa_2(\sigma_1+4\sigma_2-2\sigma_3)+4\mu_2(\sigma_1+\sigma_2+\\
&+\sigma_3)\Big)\Big)+\mu_2\Big(\kappa_1\Big(-9\kappa_2^2\delta\mu(\sigma_1-2\sigma_2+\sigma_3)+6\kappa_2\mu_2\Big(\mu_1\Big(\sigma_1-2(\sigma_2+\sigma_3)\Big)+\mu_2(-2\sigma_1+\\
&+\sigma_2+\sigma_3)\Big)-8\mu_1\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\mu_1\Big(9\kappa_2\mu_1(\sigma_3-\sigma_2)+3\kappa_2\mu_2(\sigma_1+4\sigma_2-\\
&-2\sigma_3)+4\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)\Big),
\end{aligned}$$

and

$$\xi_3 = -\,\theta\delta\mu\Big(\kappa_1\Big(-9\kappa_2^2\delta\mu(\sigma_1+\sigma_2-2\sigma_3)+6\kappa_2\mu_2\Big(\mu_1(\sigma_1+\sigma_2-5\sigma_3)+\mu_2(-2\sigma_1+\sigma_2+ +\sigma_3)\Big)-8\mu_1\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\mu_1\mu_2\Big(3\kappa_2(\sigma_1-2\sigma_2+4\sigma_3)+4\mu_2(\sigma_1+\sigma_2+ +\sigma_3)\Big)\Big)+\mu_2\Big(\kappa_1\Big(-9\kappa_2^2\delta\mu(\sigma_1+\sigma_2-2\sigma_3)+6\kappa_2\mu_2\Big(\mu_1\Big(\sigma_1-2(\sigma_2+\sigma_3)\Big)+\mu_2(-2\sigma_1+ +\sigma_2+\sigma_3)\Big)-8\mu_1\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\mu_1\Big(9\kappa_2\mu_1(\sigma_2-\sigma_3)+3\kappa_2\mu_2(\sigma_1-2\sigma_2+4\sigma_3)+ +4\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)\Big).$$

The maximum value of the function F in this case reads

$$F(\eta_1,\eta_2,\eta_3) = \frac{4f_3(\sigma_1,\sigma_2,\sigma_3)}{(\theta\delta\mu-\mu_2)\zeta_3}.$$

Condition $\eta_1-\eta_2\le 0$ is equivalent to

$$-6\theta^2\delta\kappa\delta\mu\mu_2\sigma_1+\mu_2\Big(-6\kappa_2\mu_2\sigma_1+3\kappa_1\kappa_2(-\sigma_1+\sigma_2)+2\kappa_2\mu_1(\sigma_2-\sigma_3) +2\kappa_1\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)+\theta\Big(2\kappa_2\mu_2\Big(6\mu_2\sigma_1+\mu_1(-3\sigma_1-\sigma_2+\sigma_3)\Big) +\kappa_1\Big(3\kappa_2\delta\mu(\sigma_1-\sigma_2)+2\mu_1\mu_2(\sigma_1+2\sigma_2)-2\mu_2^2(4\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\ge 0,$$

condition $\eta_2-\eta_3\le 0$ is equivalent to $\sigma_2\le\sigma_3$, which is the assumption of the theorem, and condition $-\mu_2(\eta_1+\eta_3)-\lambda_2(\eta_1+\eta_2+\eta_3)\le 0$ is equivalent to

$$-6\theta^2\delta\kappa\delta\mu^2(3\kappa_2+\mu_2)\sigma_1-\theta\delta\mu\Big(\kappa_1\kappa_2(-9\mu_1\sigma_1+33\mu_2\sigma_1+6\mu_2\sigma_2+9\mu_1\sigma_3-3\mu_2\sigma_3)- -6\kappa_2^2\Big(6\mu_2\sigma_1+\mu_1(-2\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_1\mu_2\Big(\mu_2(4\sigma_1+\sigma_2+\sigma_3)+3\mu_1(\sigma_1+2\sigma_3)\Big)- -2\kappa_2\mu_2\Big(6\mu_2\sigma_1+\mu_1(\sigma_1+\sigma_2+7\sigma_3)\Big)\Big)+\mu_2\Big(\kappa_1\Big(3\kappa_2\mu_2(5\sigma_1+2\sigma_2-\sigma_3)+9\kappa_2\mu_1(-\sigma_1+ +\sigma_3)+2\mu_2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)+2\kappa_2\Big(-3\mu_2^2\sigma_1+3\mu_1^2(-\sigma_2+\sigma_3)-\mu_1\mu_2(\sigma_1+ +\sigma_2+7\sigma_3)-3\kappa_2\Big(3\mu_2\sigma_1+\mu_1(-2\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\Big)\le 0,$$

which gives case H, and concludes the first part of the proof.

II. Let us now consider our function on the part where $(\mu_2+\lambda_2)(\eta_1+\eta_3)+\lambda_2\eta_2\le 0$. This implies that the function F is of the form

$$F(\eta_1,\eta_2,\eta_3) = 2(\sigma_1\eta_1+\sigma_2\eta_2+\sigma_3\eta_3)-\frac{2\mu_1\mu_2}{\delta\mu}(\eta_1^2+\eta_2^2+\eta_3^2)- -\left(\frac{\kappa_1\kappa_2}{\delta\kappa}-\frac{2\mu_1\mu_2}{3\delta\mu}\right)(\eta_1+\eta_2+\eta_3)^2-(1-\theta)\left(2\mu_2(\eta_1^2+\eta_2^2+\eta_3^2)+\lambda_2(\eta_1+\eta_2+\eta_3)^2- -\frac{1}{2\mu_2+\lambda_2}\Big(2\mu_2\eta_3+\lambda_2(\eta_1+\eta_2+\eta_3)\Big)^2\right),$$

and we shall maximize it over $\mathbf{R}^3$, under constrains

$$\begin{aligned} \eta_1 - \eta_2 &\leq 0, \\ \eta_2 - \eta_3 &\leq 0, \\ (\mu_2 + \lambda_2)(\eta_1 + \eta_3)\lambda_2\eta_2 &\leq 0. \end{aligned}$$

The KKT system in this case reads

$$\begin{cases} -2\sigma_1 + \dfrac{4\mu_1\mu_2}{\delta\mu}\eta_1 + 2\alpha(\eta_1+\eta_2+\eta_3) + (1-\theta)(4\mu_2\eta_1 + \beta(\eta_2+\eta_3)) + a_1 + a_3(\mu_2+\lambda_2) = 0, \\ -2\sigma_2 + \dfrac{4\mu_1\mu_2}{\delta\mu}\eta_2 + 2\alpha(\eta_1+\eta_2+\eta_3) + (1-\theta)(4\mu_2\eta_2 + \beta(\eta_2+\eta_3)) - a_1 + a_2 + a_3\lambda_2 = 0, \\ -2\sigma_3 + \dfrac{4\mu_1\mu_2}{\delta\mu}\eta_3 + 2\alpha(\eta_1+\eta_2+\eta_3) - a_2 + a_3(\mu_2+\lambda_2) = 0, \\ a_1(\eta_1 - \eta_2) = 0, \\ a_2(\eta_2 - \eta_3) = 0, \\ a_3((\mu_2+\lambda_2)(\eta_1+\eta_3) + \lambda_2\eta_2) = 0, \\ \eta_1 - \eta_2 \leq 0, \\ \eta_2 - \eta_3 \leq 0, \\ (\mu_2+\lambda_2)(\eta_1+\eta_3) + \lambda_2\eta_2 \leq 0, \\ a_1, a_2, a_3 \geq 0, \end{cases}$$

where $\alpha = \frac{\kappa_1\kappa_2}{\delta\kappa} - \frac{2\mu_1\mu_2}{3\delta\mu}$, and $\beta = \frac{4\mu_2\lambda_2}{2\mu_2+\lambda_2}$. Again, there are three complementary conditions, and we need to check 8 cases.

Case 1. If $\eta_1 = \eta_2 = \eta_3$ and $(\mu_2+\lambda_2)(\eta_1+\eta_3) + \lambda_2\eta_2 = 0$, then the optimal is $\eta_1 = \eta_2 = \eta_3 = 0$, and therefore $F(\eta_1, \eta_2, \eta_3) = 0$. We calculate parameters

$$a_1 = \frac{6\kappa_2(2\sigma_1 - \sigma_2 - \sigma_3) - 2\mu_2(\sigma_1+\sigma_2+\sigma_3)}{9\kappa_2},$$

$$a_2 = \frac{2\Big(3\kappa_2(\sigma_1+\sigma_2-2\sigma_3) + \mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)}{9\kappa_2}, \ a_3 = \frac{2(\sigma_1+\sigma_2+\sigma_3)}{3\kappa_2}.$$

Since $\kappa_2 > 0$, nonnegativity of parameters a_1, a_2 and a_3 implies

$$\begin{aligned} \sigma_1+\sigma_2+\sigma_3 &\geq 0 \\ 3\kappa_2(\sigma_1+\sigma_2-2\sigma_3) + \mu_2(\sigma_1+\sigma_2+\sigma_3) &\geq 0 \\ 3\kappa_2(2\sigma_1-\sigma_2-\sigma_3) - \mu_2(\sigma_1+\sigma_2+\sigma_3) &\geq 0, \end{aligned}$$

which, together with case I.1, concludes case A of the theorem.

Case 2. If $\eta_1 = \eta_2 = \eta_3$ and $a_3 = 0$, then

$$\eta_1 = \frac{\delta\kappa(3\kappa_2 + 4\mu_2)(\sigma_1 + \sigma_2 + \sigma_3)}{9\kappa_2\Big(4(1-\theta)\kappa_2\mu_2 + \kappa_1(3\kappa_2 + 4\theta\mu_2)\Big)},$$

$$a_1 = \frac{12(1-\theta)\kappa_2\mu_2(\sigma_1 - \sigma_2 - \sigma_3) + 2\kappa_1(\kappa_2(6\sigma_1 - 3(\sigma_2 + \sigma_3))}{3\Big(\kappa_1(4\theta\mu_2 + 3\kappa_2) + 4(1-\theta)\kappa_2\mu_2\Big)} + \\ + \frac{2\mu_2(\sigma_1 + 3\theta\sigma_1 + (1-3\theta)(\sigma_2 + \sigma_3))}{3\Big(\kappa_1(4\theta\mu_2 + 3\kappa_2) + 4(1-\theta)\kappa_2\mu_2\Big)},$$

and

$$a_2 = \frac{8\kappa_1\mu_2(-3\theta\sigma_3 + \sigma_1 + \sigma_2 + \sigma_3) + 24(\theta - 1)\kappa_2\mu_2\sigma_3 + 6\kappa_1\kappa_2(\sigma_1 + \sigma_2 - 2\sigma_3)}{3\Big(\kappa_1(4\theta\mu_2 + 3\kappa_2) + 4(1-\theta)\kappa_2\mu_2\Big)}.$$

The maximum value of the function F in this case is

$$F(\eta_1, \eta_2, \eta_3) = \frac{\delta\kappa(3\kappa_2 + 4\mu_2)(\sigma_1 + \sigma_2 + \sigma_3)^2}{9\kappa_2\Big(4(1-\theta)\kappa_2\mu_2 + \kappa_1(3\kappa_2 + 4\theta\mu_2)\Big)}.$$

Condition $(\mu_2 + \lambda_2)(\eta_1 + \eta_3) + \lambda_2\eta_2 \leq 0$ is equivalent to

$$\sigma_1 + \sigma_2 + \sigma_3 \leq 0,$$

and together with $a_1 \geq 0$ and $a_2 \geq 0$ we obtain (3.49), which concludes case B of the theorem.

Case 3. Assume that $\eta_1 = \eta_2$, $(\mu_2 + \lambda_2)(\eta_1 + \eta_3) + \lambda_2\eta_2 = 0$ and $a_2 = 0$. Then

$$\eta_1 = \frac{-\delta\kappa\delta\mu(3\kappa_2 + \mu_2)(3\kappa_2 + 4\mu_2)}{\zeta_1}\Big(3\kappa_2(\sigma_1 + \sigma_2 - 2\sigma_3) + \mu_2(\sigma_1 + \sigma_2 + \sigma_3)\Big),$$

$$\begin{aligned} a_1 = \frac{9\kappa_2\mu_2}{\zeta_1}\Big(&4\theta\delta\kappa\delta\mu(3\kappa_2 + \mu_2)\Big((3\kappa_2 + \mu_2)(\sigma_1 - \sigma_2) - 3\mu_2\sigma_3\Big) - 2\mu_2\Big(\kappa_1\Big(18\kappa_2^2(\sigma_2 - \sigma_1) - \\ &- 9\kappa_2\mu_1(\sigma_1 - 2\sigma_2 + \sigma_3) + 3\kappa_2\mu_2(-3\sigma_1 + 4\sigma_2 + 5\sigma_3) + 2\mu_2(3\mu_1 + \mu_2)(\sigma_1 + \sigma_2 + \sigma_3)\Big) + \\ &+ 2\kappa_2\Big(9\kappa_2^2(\sigma_1 - \sigma_2) + 3\kappa_2\Big(\mu_1(\sigma_1 - 3\sigma_2 + 2\sigma_3) + \mu_2(2\sigma_1 - 2\sigma_2 - 3\sigma_3)\Big) + \\ &+ \mu_2\Big(\mu_2(\sigma_1 - \sigma_2 - 3\sigma_3) - \mu_1(5\sigma_1 + 3\sigma_2 + \sigma_3)\Big)\Big)\Big)\Big), \end{aligned}$$

and

$$\begin{aligned} a_3 = \frac{9\kappa_2\mu_2}{\zeta_1}\Big(&6\kappa_1\kappa_2\Big(3\mu_1(4\theta\sigma_3 + \sigma_1 + \sigma_2 - 2\sigma_3) + \mu_2\Big(2(5 - 6\theta)\sigma_3 + \sigma_1 + \sigma_2\Big)\Big) + \\ &+ 8\kappa_1\mu_2\Big(3\mu_1(\theta\sigma_3 + \sigma_1 + \sigma_2 + \sigma_3) + \mu_2(-3\theta\sigma_3 + \sigma_1 + \sigma_2 + \sigma_3)\Big) - 8\kappa_2\Big(3\kappa_2\mu_1\Big((3\theta - \\ &- 2)\sigma_3 + \sigma_1 + \sigma_2\Big) + 9(1-\theta)\kappa_2\mu_2\sigma_3 + \mu_2\sigma_3\Big((1 + 3\theta)\mu_1 + 3(1-\theta)\mu_2\Big) + 4\mu_1\mu_2(\sigma_1 + \sigma_2)\Big)\Big). \end{aligned}$$

By introducing η_1, $\eta_2 = \eta_2 =$ and $\eta_3 = -\frac{\mu_2+2\lambda_2}{\mu_2+\lambda_2}\eta_1$ into the function F, it follows

$$F(\eta_1,\eta_2,\eta_3) = \frac{-\delta\kappa\delta\mu(3\kappa_2+4\mu_2)}{\zeta_1}\Big(3\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)^2,$$

which corresponds to bound (3.54). The condition $\eta_2 - \eta_3 \leq 0$ is equivalent to

$$3\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3) \leq 0,$$

which together with $a_1 \geq 0$ gives (3.51), while (3.53) is obtained from $a_3 \geq 0$.

Case 4. Let $\eta_1 = \eta_2$ and $a_2 = a_3 = 0$. Then

$$\eta_1 = \frac{-(3\kappa_2+4\mu_2)}{\zeta_1}\Big(3\kappa_1\kappa_2\delta\mu(\sigma_1+\sigma_2-2\sigma_3)+4\delta\kappa\mu_1\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big),$$

$$\eta_3 = \frac{2\xi_3}{\zeta_1}, \quad \text{and} \quad a_1 = \sigma_1 - \sigma_2,$$

where

$$\begin{aligned}\xi_3 =& 2\sigma_3\Big(\mu_2^2\Big(3\kappa_2\Big((5-9\theta)\kappa_1+9(\theta-1)\kappa_2\Big)-4\mu_1\delta\kappa\Big)-3\kappa_2\mu_2\Big((4-9\theta)\kappa_1\mu_1+(9\theta-\\ &-8)\kappa_2\mu_1+3\kappa_1\kappa_2\Big)+9\kappa_1\kappa_2^2\mu_1\Big)+(3\kappa_2+4\mu_2)(\sigma_1+\sigma_2)(3\kappa_1\kappa_2\delta\mu-2\mu_1\mu_2\delta\kappa).\end{aligned}$$

The maximum value of the function F is

$$F = \frac{1}{\zeta_3} f_2(\sigma_3,\sigma_2,\sigma_1).$$

Conditions $\eta_2 - \eta_3 \leq 0$ and $\mu_2(\eta_1+\eta_3)+\lambda_2(\eta_1+\eta_2+\eta_3) \leq 0$ are equivalent to

$$\begin{gathered}3\kappa_1\kappa_2(\sigma_1+\sigma_2-2\sigma_3)+12(-1+\theta)\kappa_2\mu_2\sigma_3+4\kappa_1\mu_2(\sigma_1+\sigma_2+\sigma_3-3\theta\sigma_3) \leq 0,\\ (3\kappa_2+4\mu_2)(-4\kappa_2\mu_1+\kappa_1(3\mu_1+\mu_2))(\sigma_1+\sigma_2)+2\Big(3\kappa_2\Big((-3+6\theta)\kappa_1+\\ (4-6\theta)\kappa_2\Big)\mu_1+\Big(3\kappa_2\Big((5-6\theta)\kappa_1+6(-1+\theta)\kappa_2\Big)-2\Big(-3(1+\theta)\kappa_1+\kappa_2+\\ 3\theta\kappa_2\Big)\mu_1\Big)\mu_2+2\Big(\kappa_1-3\theta\kappa_1+3(-1+\theta)\kappa_2\Big)\mu_2^2\Big)\sigma_3 \geq 0,\end{gathered}$$

which, together with $a_1 \geq 0$, correspond to (3.69) and case I of Theorem 3.45.

Case 5. If $a_1 = 0$, $\eta_2 = \eta_3$ and $(\mu_2+\lambda_2)(\eta_1+\eta_3)+\lambda_2\eta_2 = 0$ then

$$\eta_2 = \frac{\delta\kappa\delta\mu(3\kappa_2+\mu_2)(3\kappa_2+4\mu_2)}{\zeta_1}\Big(\kappa_2\Big(6\sigma_1-3(\sigma_2+\sigma_3)\Big)-\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big),$$

$$a_2 = \frac{9\kappa_2\mu_2}{\zeta_1}\Big(2\kappa_1\Big(18\kappa_2^2(\sigma_2-\sigma_3)(-\theta\delta\mu+\mu_2)+3\kappa_2\mu_2\Big(\mu_1(6\theta\sigma_1+4\theta\sigma_2-4\theta\sigma_3-3\sigma_1+$$
$$+6\sigma_2-3\sigma_3)+\mu_2(-6\theta\sigma_1-4\theta\sigma_2+4\theta\sigma_3+5\sigma_1+4\sigma_2-3\sigma_3)\Big)+2\mu_2^2\Big(\mu_1\Big(3(\theta+1)\sigma_1+$$
$$+(\theta+3)\sigma_2-(\theta-3)\sigma_3\Big)+\mu_2(-3\theta\sigma_1-\theta\sigma_2+\theta\sigma_3+\sigma_1+\sigma_2+\sigma_3)\Big)\Big)-4\kappa_2\Big(9\kappa_2^2(\sigma_2-$$
$$-\sigma_3)(-\theta\delta\mu+\mu_2)-3\kappa_2\mu_2\Big(\mu_1(-3\theta\sigma_1-2\theta\sigma_2+2\theta\sigma_3+2\sigma_1-3\sigma_2+\sigma_3)-(1-$$
$$-\theta)\mu_2(3\sigma_1+2\sigma_2-2\sigma_3)\Big)+\mu_2^2\Big(\mu_1\Big(3\theta\sigma_1+(\theta+3)\sigma_2-(\theta-5)\sigma_3+\sigma_1\Big)-$$
$$-(1-\theta)\mu_2(-3\sigma_1-\sigma_2+\sigma_3)\Big)\Big)\Big)$$

and

$$a_3 = \frac{9\kappa_2\mu_2}{\zeta_1}\Big(6\kappa_1\kappa_2\Big(3(2\theta-1)\mu_1(\sigma_2+\sigma_3)-(6\theta-7)\mu_2(\sigma_2+\sigma_3)+6\mu_1\sigma_1-2\mu_2\sigma_1\Big)+$$
$$+4\kappa_1\mu_2\Big(3\mu_1\Big(3\theta\sigma_1+(\theta+1)(\sigma_2+\sigma_3)+\sigma_1\Big)+\mu_2\Big((5-9\theta)\sigma_1-(3\theta-5)(\sigma_2+\sigma_3)\Big)\Big)-$$
$$-4\kappa_2\Big(\kappa_2\Big(3(3\theta-1)\mu_1(\sigma_2+\sigma_3)-9(\theta-1)\mu_2(\sigma_2+\sigma_3)+6\mu_1\sigma_1\Big)+\mu_2\Big((9\theta-1)\mu_1\sigma_1+$$
$$+(3\theta+5)\mu_1(\sigma_2+\sigma_3)-3(\theta-1)\mu_2(3\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\Big).$$

By introducing $\eta_1 = -\dfrac{\mu_2+2\lambda_2}{\mu_2+\lambda_2}\eta_3$, η_2 and $\eta_3=\eta_2$ into the function F, we obtain

$$F(\eta_1,\eta_2,\eta_3) = \frac{-\delta\kappa\delta\mu(3\kappa_2+4\mu_2)}{\zeta_1}\Big(3\kappa_2(-2\sigma_1+\sigma_2+\sigma_3)+\mu_2(\sigma_1+\sigma_2+\sigma_3)\Big)^2,$$

which implies the bound (3.60) in case E of the theorem. The condition $\eta_1-\eta_2\le 0$ is equivalent to

$$6\kappa_2\sigma_1-3\kappa_2(\sigma_2+\sigma_3)-\mu_2(\sigma_1+\sigma_2+\sigma_3)\le 0,$$

which, together with condition $a_2\ge 0$, gives (3.57), while $a_3\ge 0$ gives the condition (3.59), which concludes case E.

Case 6. If $\eta_2=\eta_3$ and $a_1=a_3=0$, then

$$\eta_1=\frac{2\xi_1}{\zeta_2},\quad \eta_2=\frac{\xi_2}{\zeta_2}$$

where

$$\xi_1 =\kappa_2\mu_2\Big(3\kappa_2\Big((6\theta-4)\mu_1\sigma_1-(3\theta-5)\mu_1(\sigma_2+\sigma_3)-3(\theta-1)\mu_2(2\sigma_1-\sigma_2-\sigma_3)\Big)+$$
$$+2\mu_2\Big(3\theta\mu_1+3(1-\theta)\mu_2+\mu_1)(\sigma_1+\sigma_2+\sigma_3)\Big)-\kappa_1\Big(3\kappa_2\mu_2\Big(\mu_1\Big((6\theta+4)\sigma_1-$$
$$-3\theta(\sigma_2+\sigma_3)+\sigma_2+\sigma_3\Big)-(3\theta+1)\mu_2\Big(2\sigma_1-\sigma_2-\sigma_3\Big)\Big)+2\mu_2^2\Big(3\theta\mu_1-3(\theta-1)\mu_2+$$
$$+\mu_1\Big)(\sigma_1+\sigma_2+\sigma_3)-9\kappa_2^2\delta\mu(2\sigma_1-\sigma_2-\sigma_3)\Big),$$

and

$$\begin{aligned}\xi_2 =&\kappa_1\Big(- 6\kappa_2\mu_2\Big(\mu_2\Big(3\theta\sigma_1 - 2(3\theta-2)(\sigma_2+\sigma_3)+\sigma_1\Big) - (3\theta-1)\mu_1\Big(\sigma_1 - 2(\sigma_2+\sigma_3)\Big)\Big)-\\ &-4\mu_2^2\Big(3\theta\mu_1 - 3(\theta-1)\mu_2+\mu_1\Big)(\sigma_1+\sigma_2+\sigma_3) - 9\kappa_2^2\delta\mu(2\sigma_1-\sigma_2-\sigma_3)\Big)+\\ &+2\kappa_2\mu_2\Big(- 3\kappa_2\mu_1\Big((3\theta-5)\sigma_1 - 2(3\theta-2)(\sigma_2+\sigma_3)\Big) + 9(\theta-1)\kappa_2\mu_2\Big(\sigma_1 - 2(\sigma_2+\sigma_3)\Big)+\\ &+2\mu_2\Big(3\theta\mu_1+3(1-\theta)\mu_2+\mu_1\Big)(\sigma_1+\sigma_2+\sigma_3)\Big).\end{aligned}$$

Moreover,

$$\begin{aligned}a_2 = &\frac{36\kappa_2\mu_2}{\zeta_2}\Big(\kappa_2\Big(\mu_1\Big(3\theta\kappa_1(\sigma_2-\sigma_3)+2(\theta-1)\mu_2(3\theta\sigma_3+\sigma_1-\sigma_2)\Big)-\\ &- 3(1-\theta)\mu_2\Big(\kappa_1(\sigma_3-\sigma_2) - 2(\theta-1)\mu_2\sigma_3\Big)\Big) + 2\kappa_1\mu_2\Big(\theta\mu_1(-3\theta\sigma_3+2\sigma_2+\sigma_3)+\\ &+ (1-\theta)\mu_2(-3\theta\sigma_3+\sigma_1+\sigma_2+\sigma_3)\Big)\Big).\end{aligned}$$

The maximum value of the function F in this case is

$$F(\eta_1,\eta_2,\eta_3) = \frac{1}{\zeta_2} f_1(\sigma_3,\sigma_2,\sigma_1),$$

which gives the bound (3.72). The condition $\eta_1 - \eta_2 \le 0$ is equivalent to

$$\begin{aligned}6(1-\theta)\kappa_2\mu_2(\sigma_1-\sigma_2-\sigma_3) + \kappa_1\Big(3\kappa_2(2\sigma_1-\sigma_2-\sigma_3)+\\ +2\mu_2\Big(\sigma_1+\sigma_2+\sigma_3 - 3\theta(\sigma_2+\sigma_3-\sigma_1)\Big)\Big) \le 0,\end{aligned}$$

while the condition $\mu_2(\eta_1+\eta_3)+\lambda_2(\eta_1+\eta_2+\eta_3) \le 0$ is equivalent to

$$\begin{aligned}2\kappa_2\Big(9(-1+\theta)\kappa_2\mu_2(\sigma_2+\sigma_3)+3(-1+\theta)\mu_2^2(3\sigma_1+\sigma_2+\sigma_3)+\\ 3\kappa_2\mu_1\Big(- 2\sigma_1 - (-1+3\theta)(\sigma_2+\sigma_3)\Big) - \mu_1\mu_2\Big((-1+9\theta)\sigma_1+(5+3\theta)(\sigma_2+\sigma_3)\Big)\Big)+\\ \kappa_1\Big(3\kappa_2\Big(6\mu_1\sigma_1 - 2\mu_2\sigma_1+3(-1+2\theta)\mu_1(\sigma_2+\sigma_3) - (-7+6\theta)\mu_2(\sigma_2+\sigma_3)\Big)+\\ 2\mu_2\Big(3\mu_1\Big(\sigma_1+3\theta\sigma_1+(1+\theta)(\sigma_2+\sigma_3)\Big)+\mu_2\Big((5-9\theta)\sigma_1-(-5+3\theta)(\sigma_2+\sigma_3)\Big)\Big)\Big) \ge 0.\end{aligned}$$

Together with condition $a_2 \ge 0$, we obtain (3.71) of case J.

Case 7. Let $(\mu_2+\lambda_2)(\eta_1+\eta_3)+\lambda_2\eta_2 = 0$ and $a_1 = a_2 = 0$. Then

$$\eta_1 = \frac{\xi_1}{\zeta_4}, \quad \eta_3 = \frac{\xi_3}{\zeta_4},$$

where

$$\begin{aligned}\xi_1 =&\delta\mu\Big(2\theta\delta\kappa\delta\mu(3\kappa_2+\mu_2)\Big(9\kappa_2^2(2\sigma_1-\sigma_2-\sigma_3)+3\kappa_2\mu_2(4\sigma_1+\sigma_2-8\sigma_3)+\\ &+2\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)+\mu_2\Big(\kappa_1\Big(54\kappa_2^3(2\sigma_1-\sigma_2-\sigma_3)+27\kappa_2^2\Big(\mu_1(\sigma_1-2\sigma_2+\sigma_3)+\end{aligned}$$

$$+\mu_2(3\sigma_1-5\sigma_3)\Big)+18\kappa_2\mu_2\Big(\mu_1(-2\sigma_1+\sigma_2-2\sigma_3)+\mu_2\sigma_2\Big)+4\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+$$
$$+\sigma_3)\Big)-2\kappa_2\Big(27\kappa_2^3(2\sigma_1-\sigma_2-\sigma_3)+27\kappa_2^2(-\mu_1\sigma_2+\mu_1\sigma_3+2\mu_2\sigma_1-3\mu_2\sigma_3)+$$
$$+9\kappa_2\mu_2\Big(\mu_1(\sigma_2-4\sigma_1)+\mu_2(2\sigma_1+\sigma_2-2\sigma_3)\Big)+2\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\Big)$$

and

$$\xi_3=\delta\mu\Big(\mu_2\Big(\kappa_1\Big(-54\kappa_2^3(\sigma_1+\sigma_2-2\sigma_3)+27\kappa_2^2\Big(\mu_1(\sigma_1-2\sigma_2+\sigma_3)-5\mu_2\sigma_1+3\mu_2\sigma_3\Big)+$$
$$+18\kappa_2\mu_2\Big(\mu_1(-2\sigma_1+\sigma_2-2\sigma_3)+\mu_2\sigma_2\Big)+4\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)+$$
$$+2\kappa_2\Big(27\kappa_2^3(\sigma_1+\sigma_2-2\sigma_3)+27\kappa_2^2(-\mu_1\sigma_1+\mu_1\sigma_2+3\mu_2\sigma_1-2\mu_2\sigma_3)-9\kappa_2\mu_2\Big(\mu_1(\sigma_2-$$
$$-4\sigma_3)+\mu_2(-2\sigma_1+\sigma_2+2\sigma_3)\Big)-2\mu_2^2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)\Big)-2\theta\delta\kappa\delta\mu(3\kappa_2+$$
$$+\mu_2)\Big(9\kappa_2^2(\sigma_1+\sigma_2-2\sigma_3)+3\kappa_2\mu_2(8\sigma_1-\sigma_2-4\sigma_3)-2\mu_2^2(\sigma_1+\sigma_2+\sigma_3)\Big)\Big).$$

Moreover,

$$a_3=\frac{-36\kappa_2\mu_2}{\zeta_4}\Big(6\theta^2\sigma_3\delta\kappa(3\kappa_2+\mu_2)\delta\mu^2+\theta\delta\mu\Big(\kappa_1\Big(\kappa_2\Big(9\mu_1(\sigma_1-\sigma_3)-3\mu_2\sigma_1+$$
$$+6\mu_2\sigma_2+33\mu_2\sigma_3\Big)+2\mu_2\Big(3\mu_1(2\sigma_1+\sigma_3)+\mu_2(\sigma_1+\sigma_2+4\sigma_3)\Big)\Big)-2\kappa_2\Big(3\kappa_2\Big(\mu_1(\sigma_1+$$
$$+\sigma_2-2\sigma_3)+6\mu_2\sigma_3\Big)+\mu_2\Big(\mu_1(7\sigma_1+\sigma_2+\sigma_3)+6\mu_2\sigma_3\Big)\Big)\Big)+\mu_2\Big(2\kappa_2\Big(3\kappa_2\mu_1(\sigma_1+\sigma_2-$$
$$-2\sigma_3)+9\kappa_2\mu_2\sigma_3+3\mu_1^2(\sigma_2-\sigma_1)+\mu_1\mu_2(7\sigma_1+\sigma_2+\sigma_3)+3\mu_2^2\sigma_3\Big)-\kappa_1\Big(9\kappa_2\mu_1(\sigma_1-$$
$$-\sigma_3)-3\kappa_2\mu_2(\sigma_1-2\sigma_2-5\sigma_3)+2\mu_2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)\Big)\Big)\Big).$$

Putting η_1, $\eta_2=-\frac{\mu_2+\lambda_2}{\lambda_2}(\eta_1+\eta_3)$ and η_3 into the function F, we get

$$F(\eta_1,\eta_2,\eta_3)=\frac{1}{\zeta_4}f_4(\sigma_1,\sigma_2,\sigma_3).$$

The condition $\eta_1-\eta_2\le 0$ is equivalent to

$$-2\theta\delta\kappa\delta\mu(3\kappa_2+\mu_2)\Big((3\kappa_2+\mu_2)(\sigma_1-\sigma_2)-3\mu_2\sigma_3\Big)+\mu_2\Big(2\kappa_2\Big(9\kappa_2^2(\sigma_1-\sigma_2)+$$
$$+\mu_2\Big(\mu_2(\sigma_1-\sigma_2-3\sigma_3)-\mu_1(5\sigma_1+3\sigma_2+\sigma_3)\Big)+3\kappa_2\Big(\mu_2(2\sigma_1-2\sigma_2-3\sigma_3)+$$
$$+\mu_1(\sigma_1-3\sigma_2+2\sigma_3)\Big)\Big)+\kappa_1\Big(18\kappa_2^2(-\sigma_1+\sigma_2)+2\mu_2(3\mu_1+\mu_2)(\sigma_1+\sigma_2+\sigma_3)+$$
$$+3\kappa_2\Big(-3\mu_1(\sigma_1-2\sigma_2+\sigma_3)+\mu_2(-3\sigma_1+4\sigma_2+5\sigma_3)\Big)\Big)\Big)\le 0,$$

while the condition $\eta_2 - \eta_3 \leq 0$ is equivalent to

$$-2\kappa_2\Big(9\kappa_2^2(-\theta\delta\mu + \mu_2)(\sigma_2 - \sigma_3) + \mu_2^2\Big((1-\theta)\mu_2(3\sigma_1 + \sigma_2 - \sigma_3) + \mu_1\Big(\sigma_1 + 3\theta\sigma_1 + \\ +(3+\theta)\sigma_2 + (5-\theta)\sigma_3\Big)\Big) - 3\kappa_2\mu_2\Big((-1+\theta)\mu_2(3\sigma_1 + 2\sigma_2 - 2\sigma_3) + \mu_1(2\sigma_1 - 3\theta\sigma_1 - 3\sigma_2 - \\ -2\theta\sigma_2 + \sigma_3 + 2\theta\sigma_3)\Big)\Big) + \kappa_1\Big(18\kappa_2^2\Big(\theta(\mu_1 - \mu_2) + \mu_2)(\sigma_2 - \sigma_3) + 2\mu_2^2\Big(\mu_1\Big(3(1+\theta)\sigma_1 + (3+ \\ +\theta)\sigma_2 - (-3+\theta)\sigma_3\Big) + \mu_2(\sigma_1 - 3\theta\sigma_1 + \sigma_2 - \theta\sigma_2 + \sigma_3 + \theta\sigma_3)\Big) + 3\kappa_2\mu_2\Big(\mu_1(-3\sigma_1 + \\ +6\theta\sigma_1 + 6\sigma_2 + 4\theta\sigma_2 - 3\sigma_3 - 4\theta\sigma_3) + \mu_2(5\sigma_1 - 6\theta\sigma_1 + 4\sigma_2 - 4\theta\sigma_2 - 3\sigma_3 + 4\theta\sigma_3)\Big)\Big) \geq 0.$$

The above conditions and nonnegativity of a_3 give (3.63) and (3.65), and value of the function F in this case gives the bound (3.66), which concludes case G of the theorem.

Case 8. If $a_1 = a_2 = a_3 = 0$, then

$$\eta_1 = \frac{2\xi_1}{(\theta\delta\mu - \mu_2)\zeta_3}, \quad \eta_2 = \frac{2\xi_2}{(\theta\delta\mu - \mu_2)\zeta_3}, \quad \eta_3 = \frac{2\xi_3}{\zeta_3},$$

where

$$\xi_1 = \mu_2\Big(\kappa_1\Big(9\kappa_2^2\delta\mu(2\sigma_1 - \sigma_2 - \sigma_3) + 6\kappa_2\mu_2\Big((\mu_2 - 2\mu_1)(\sigma_1 + \sigma_2) + \sigma_3(\mu_1 - 2\mu_2)\Big) - \\ -8\mu_1\mu_2^2(\sigma_1 + \sigma_2 + \sigma_3)\Big) + 2\kappa_2\mu_1\Big(9\kappa_2\mu_1(\sigma_2 - \sigma_1) + 3\kappa_2\mu_2(4\sigma_1 - 2\sigma_2 + \sigma_3) + \\ +4\mu_2^2(\sigma_1 + \sigma_2 + \sigma_3)\Big)\Big) + \theta\delta\mu\Big((-3\kappa_1\kappa_2\delta\mu - 4\mu_1\mu_2\delta\kappa)(\kappa_2(6\sigma_1 - 3\sigma_2) + 2\mu_2(\sigma_1 + \sigma_2)\Big) + \\ +\sigma_3(3\kappa_2 + 4\mu_2)(3\kappa_1\kappa_2\delta\mu - 2\mu_1\mu_2\delta\kappa)\Big),$$

$$\xi_2 = -\theta\delta\mu\Big((-3\kappa_1\kappa_2\delta\mu + 4\kappa_1\mu_1\mu_2 - 4\kappa_2\mu_1\mu_2)\Big(3\kappa_2(\sigma_1 - 2\sigma_2) - 2\mu_2(\sigma_1 + \sigma_2)\Big) + \\ +\sigma_3(3\kappa_2 + 4\mu_2)(-3\kappa_1\kappa_2\delta\mu + 2\mu_1\mu_2\delta\kappa)\Big) + \mu_2\Big(\kappa_1\Big(- 9\kappa_2^2\delta\mu(\sigma_1 - 2\sigma_2 + \sigma_3) + \\ +6\kappa_2\mu_2\Big((\mu_2 - 2\mu_1)(\sigma_1 + \sigma_2) + \sigma_3(\mu_1 - 2\mu_2)\Big) - 8\mu_1\mu_2^2(\sigma_1 + \sigma_2 + \sigma_3)\Big) + \\ +2\kappa_2\mu_1\Big(9\kappa_2\mu_1(\sigma_1 - \sigma_2) + 3\kappa_2\mu_2(-2\sigma_1 + 4\sigma_2 + \sigma_3) + 4\mu_2^2(\sigma_1 + \sigma_2 + \sigma_3)\Big)\Big),$$

and

$$\xi_3 = 2\sigma_3\Big(\mu_2^2\Big(3\kappa_2\Big((5-9\theta)\kappa_1 - 9(1-\theta)\kappa_2\Big) - 4\mu_1\delta\kappa\Big) - 3\kappa_2\mu_2\Big((4-9\theta)\kappa_1\mu_1 + \\ +(9\theta - 8)\kappa_2\mu_1 + 3\kappa_1\kappa_2\Big) + 9\kappa_1\kappa_2^2\mu_1\Big) + (3\kappa_2 + 4\mu_2)(\sigma_1 + \sigma_2)(3\kappa_1\kappa_2\delta\mu - 2\mu_1\mu_2\delta\kappa).$$

The maximum value of the function F is

$$F(\eta_1, \eta_2, \eta_3) = \frac{4}{(\theta\delta\mu - \mu_2)} f_3(\sigma_3, \sigma_2, \sigma_1).$$

The condition $\eta_1 - \eta_2 \leq 0$ is equivalent to $\sigma_1 \leq \sigma_2$, which is the assumption of the theorem, while from the equivalence of $\eta_2 - \eta_3 \leq 0$ with

$$\begin{aligned}2\kappa_1\mu_2\Big((1-\theta)\mu_2(\sigma_1+\sigma_2+\sigma_3-3\theta\sigma_3)+\theta\mu_1(2\sigma_2+\sigma_3-3\theta\sigma_3)\Big)+\\+\kappa_2\Big(-3(1-\theta)\mu_2\Big(2(1-\theta)\mu_2\sigma_3+\kappa_1(-\sigma_2+\sigma_3)\Big)+\mu_1\Big(3\theta\kappa_1(\sigma_2-\sigma_3)-\\-2(1-\theta)\mu_2(\sigma_1-\sigma_2+3\theta\sigma_3)\Big)\Big)\leq 0\end{aligned}$$

and $\mu_2(\eta_1+\eta_3)+\lambda_2(\eta_1+\eta_2+\eta_3) \leq 0$ with

$$\begin{aligned}-6\theta^2\delta\kappa\delta\mu^2(3\kappa_2+\mu_2)\sigma_3-\theta\delta\mu\Big(-2\kappa_2\Big(3\kappa_2\Big(\mu_1(\sigma_1+\sigma_2-2\sigma_3)+6\mu_2\sigma_3\Big)+\mu_2\Big(6\mu_2\sigma_3+\\+\mu_1(7\sigma_1+\sigma_2+\sigma_3)\Big)\Big)+\kappa_1\Big(\kappa_2(9\mu_1\sigma_1-3\mu_2\sigma_1+6\mu_2\sigma_2-9\mu_1\sigma_3+33\mu_2\sigma_3)+\\+2\mu_2\Big(3\mu_1(2\sigma_1+\sigma_3)+\mu_2(\sigma_1+\sigma_2+4\sigma_3)\Big)\Big)\Big)+\mu_2\Big(-2\kappa_2\Big(3\mu_1^2(-\sigma_1+\sigma_2)+3\mu_2^2\sigma_3+\\+\mu_1\mu_2(7\sigma_1+\sigma_2+\sigma_3)+3\kappa_2\Big(\mu_1(\sigma_1+\sigma_2-2\sigma_3)+3\mu_2\sigma_3\Big)\Big)+\kappa_1\Big(2\mu_2(3\mu_1+\mu_2)(\sigma_1+\\+\sigma_2+\sigma_3)+3\kappa_2\Big(3\mu_1(\sigma_1-\sigma_3)+\mu_2(-\sigma_1+2\sigma_2+5\sigma_3)\Big)\Big)\Big)\geq 0\end{aligned}$$

we obtain condition (3.73). Using the value of the function F obtained in this case, we get the bound (3.74). □

Remark 3.3. An explicit calculation of the lower Hashin-Shtrikman bound on complementary energy in three space dimensions is a work in progress. It remains to carry out more tests, and try to write down these intimidating terms in a better form.

We shall also deal with the optimal microstructure for this Hashin-Shtrikman bound. As it was shown in Proposition 1.26, the bound is achieved by the sequential laminate of a rank at most 3, with the lamination directions given by eigenvectors of the matrix $\boldsymbol{\sigma}$. It remains to reveal the rank of the laminate and calculate lamination parameters.

Bibliography

Allaire, G. (1994), ‘Explicit lamination parameters for three-dimensional shape optimization’, *Control and Cybernetics* **23**, 309–326.

Allaire, G. (2002), *Shape Optimization by the Homogenization Method*, Springer-Verlag.

Allaire, G., Bonnetier, E., Francfort, G. & Jouve, F. (1997), ‘Shape optimization by the homogenization method’, *Numerische Mathematik* **76**, 27–68.

Allaire, G., Dapogny, C., Estevez, R., Faure, A. & Michailidis, G. (2017*a*), ‘Structural optimization under overhang constraints imposed by additive manufacturing processes: an overview of some recent results’, *Applied Mathematics and Nonlinear Science* **2**(2), 385–403.

Allaire, G., Dapogny, C., Estevez, R., Faure, A. & Michailidis, G. (2017*b*), ‘Structural optimization under overhang constraints imposed by additive manufacturing technologies’, *Journal of Computational Physics* **351**, 295–328.

Allaire, G., Geoffroy-Donders, P. & Pantz, O. (2018), ‘Topology optimization of modulated and oriented periodic microstructures by the homogenization method’, *Computers and Mathematics with Applications* .

Allaire, G. & Kohn, R. V. (1993*a*), ‘Explicit optimal bounds on the elastic energy of a two-phase composite in two space dimensions’, *Quart. Appl. Math* **51**, 675–699.

Allaire, G. & Kohn, R. V. (1993*b*), ‘Optimal design for minimum weight and compliance in plane stress using extremal microstructures’, *Europ. J. Mech. A/Solids* **12**(6), 839–878.

Antonić, N. & Vrdoljak, M. (2006), ‘Sequential laminates in multiple state optimal design problems’, *Math. Probl. Eng.2006* **68695**.

Antonić, N. & Vrdoljak, M. (2011), ‘Partial relaxation and optimal design problems’, *International Mathematical Forum* **33**(6), 1601–1611.

Aubin, J. P. & Ekeland, I. (1984), *Applied Nonlinear Analysis*, John Wiley and Sons, Inc.

Avellaneda, M. (1987), ‘Optimal bounds and microgeometries for elastic two-phase composites’, *SIAM J. Appl. Math* **47**(6), 1216–1228.

Ball, J. (1977), 'Convexity conditions and existence theorems in nonlinear elasticity', *Arch. Rat. Mech. Anal.* **63**, 337–403.

Bangerth, W., Hartmann, R., Kanschat, G. & al. (2000), 'deal.ii- an open source finite element library'.
URL: *https://www.dealii.org*

Bendsøe, M. (1995), *Optimization of structural topology, shape, and material*, Springer.

Bendsøe, M. & Sigmund, O. (2003), *Topology Optimization. Theory, Methods and Applications*, Springer.

Brezis, H. (2010), *Functional analysis, Sobolev spaces and partial differential equations*, Springer.

Burazin, K. (2018), 'On unique solution of multiple-state optimal design problems on annulus', *Journal of Optimization Theory and Applications* **177**, 329–344.

Burazin, K., Crnjac, I. & Vrdoljak, M. (2018), 'Variant of optimality criteria method for multiple state optimal design problems', *Communications in Mathematical Sciences* **16**(6), 1597–1614.

Burazin, K. & Vrdoljak, M. (2018), 'Exact solutions in optimal design problems for stationary diffusion equation', *Acta Applicandae Mathematica* pp. 1–18.

Casado-Díaz, J. (2015*a*), 'Smoothness properties for the optimal mixture of two isotropic materials: the compliance and eigenvalue problems', *SIAM J. Control Optim.* **53**, 2319–2349.

Casado-Díaz, J. (2015*b*), 'Some smoothness results for the optimal design of a two-composite material which minimizes the energy', *Calculus of Variations and Partial Differential Equations* **53**, 649–673.

Casado-Díaz, J., Castro, C., Luna-Laynez, M. & Zuazua, E. (2011), 'Numerical approximation of a one-dimensional elliptic optimal design problem', *Multiscale Model. Simul.* **9**, 1181–1216.

Cioranescu, D. & Donato, P. (1999), *An Introduction to Homogenization*, Oxford University Press Inc.

Dacorogna, B. (2008), *Direct Methods in the Calculus of Variations*, Springer.

Dudley, R. M. & Norvaiša, R. (2011), *Concrete Functional Calculus*, Springer.

Francfort, G. & Murat, F. (1986), 'Homogenization and optimal bounds in linearized elasticity', *The Archive for Rational Mechanics and Analysis* **94**, 307–334.

Geoffroy-Donders, P., Allaire, G., Cortial, J. & Pantz, O. (2018), 'Optimization of oriented and parametric cellular structures by the homogenization method', *Schumacer, A. et al., eds. Advances in Structural and Multidisciplinary Optimization, Springer* pp. 767–778.

Gibianski, L. & Cherkaev, A. (1984), 'Design of composite plates of extremal rigidity', *Ioffe Physicotechnical Institute preprint* .

Gibianski, L. & Cherkaev, A. (1987), 'Microstructures of composites of extremal rigidity and exact bounds of the associated energy density', *Ioffe Physicotechnical Institute preprint* .

Glowinski, R. (1984), 'Numerical simulation for some applied problems originating from continuum mechanics, in trends and applications of pure mathematics to mechanics', *P.G. Ciarlet and M. Roseau, eds., Lecture Notes in Phys. 195, Springer, Berlin* pp. 96–145.

Goodman, J., Kohn, R. & Reyna, L. (1986), 'Numerical study of a relaxed variational problem from optimal design', *Computer Methods in Applied Mechanics and Engineering* **57**, 107–127.

Grabovsky, Y. (1996), 'Bounds and extremal microstructures for two-component composites: a unified treatment based on the translation method', *Proc. Roy. Soc. London Series A* **452**, 919–944.

Hashin, Z. & Shtrikman, S. (1963), 'A variational approach to the theory of the elastic behavior of multiphase materials', *J. Mech. Phys. Solids* **11**, 127–140.

Karush, W. (1939), *Minima of Functions of Several Variables with Inequalities as Side Constraints*, M. Sc. Dissertation, Dept. of Mathematics, Univ. of Chicago.

Kirsch, U. (1981), *Optimum structural design*, McGraw-Hill, New York.

Kohn, R. V. & Lipton, R. (1988), 'Optimal bounds for the effective energy of a mixture of isotropic, incompressible, elastic materials', *Archives of Rational Mechanics and Analysis* **102**, 331–350.

Kuhn, H. W. & Tucker, A. W. (1951), 'Nonlinear programming', *Proceedings of 2nd Berkeley Symposium* pp. 481–492.

Lurie, K. & Cherkaev, A. (1984), 'Exact estimates of conductivity of composites formed by two isotropically conducting media, taken in prescribed proportion', *Proceedings of the Royal Society of Edinburgh Section A* **99**, 71–87.

Lurie, K., Cherkaev, A. & Fedorov, A. (1982), 'Regularization of Optimal Design Problems for Bars and Plates I,II', *Optim. Th. Appl.* **37**, 499–521,523–543.

Milton, G. W. & Kohn, R. V. (1988), 'Variational bounds on the effective moduli of anisotropic composites', *Journal of the Mechanics and Physics of Solids* **36**, 597–629.

Murat, F. (1971), 'Un contre-exemple pour le probléme du contrôle dans les cofficients', *C. R. Acad. Sci. Paris* **A-273**, 708–711.

Murat, F. & Tartar, L. (1978), 'H-convergence, in Séminaire d'Analyse Fonctionnelle et Numérique de l'Université d'Alger', Lecture Notes.

Murat, F. & Tartar, L. (1985), ' Calcul des Variations et Homogénéisation, in Les Méthodes de l'Homogenisation Théorie et Applications en Physique ', *Coll. Dir. Etudes et Recherches EDF, Eyrolles, Paris* **57**, 319–369.

Rockafellar, R. T. (1972), *Convex Analysis*, Princeton University Press.

Rozvany, G. I. N. (1989), *Structural design via optimality criteria*, Kluwer Academic Publishers.

Schirotzek, W. (2007), *Nonsmooth analysis*, Springer.

Spagnolo, S. (1976), 'Convergence in energy for elliptic operators', *Numerical solutions of partial differential equations III Synspade 1975, B Hubbard ed., Academic Press New York* pp. 469–498.

Tartar, L. (1975), ' Problèmes de contrôle des coefficients dans des équations aux dérivées partielles, in Control Theory, Numerical Methods and Computer Systems Modelling', *in A. Bensoussan, J.-L. Lions (Eds.), Lecture Notes in Economy and Mathematical Systems, Springer* **107**, 420–426.

Tartar, L. (1985), ' Estimations fines des coefficients homogénéisés, Ennio de Giorgi colloquium ', *P. Krée ed, Research notes in mathematics, Pitman, London,* **125**, 168–187.

Tartar, L. (1986), 'Remarks on homogenization', *Ericsen, J. L. et al., Homogenization and Effective Moduli of Materials and Media* pp. 228– 246.

Tartar, L. (1987), 'The appearance of oscillations in optimization problems', *in Non-Classical Continuum Mechanics (Durham, 1986), London Math. Soc. Lecture Note Ser 122, Cambridge University Press* pp. 129–150.

Tartar, L. (1995), 'Remarks on homogenization method in optimal design problems, in homogenization and applications to material sciences', *GAKUTO Internat. Ser. Math. Sci. Appl.* **9**, 393– 412.

Tartar, L. (2000), ‘ An introduction to the homogenization method in optimal design, in Optimal shape design’, *A. Cellina and A. Ornelas eds., Lecture Notes in Math.* **1740**, 47–156.

Toader, A.-M. (1997), ‘Convergence of an algorithm in optimal design’, *Structural Optimization* **13**, 195–198.

Vigdergauz, S. (1994), ‘Three-dimensional grained composites of exreme thermal properites’, *J. Mech. Phys. Solids* **42**, 729–740.

von Neumann, J. (1937), ‘Some matrix inequalities and metrization of matrix space’, *Tomsk University Rev.* **1**, 286–300.

Vrdoljak, M. (2010), ‘On hashin-shtrikman bounds for mixtures of two isotropic materials’, *Nonlinear Analysis: Real World Applications* **11**, 4597–4606.

Vrdoljak, M. (2016), ‘Classical optimal design in two-phase conductivity problems’, *SIAM Journal on Control and Optimization* **54**, 2020–2035.

Zhikov, V., Kozlov, S. & Oleinik, O. (1994), *Homogenization of Differential Operators and Integral Functionals*, Springer-Verlag.

Zhikov, V. V., Kozlov, S. M., Oleinik, O. & Ngoan, H. T. (1979), ‘Averaging and G-convergence of differential operators’, *Russian Math. Surveys* **34**(5), 69–147.

Curriculum vitae

Ivana Crnjac was born on September 29, 1989 in Osijek, Croatia. After finishing high school in Osijek, she enrolled in the undergraduate program in Mathematics at the Department of Mathematics, J. J. Strossmayer University of Osijek. In October 2013, she obtained a master's degree in Financial and Business Mathematics at the Department of Mathematics, University of Osijek, with master thesis 'Slaba rješenja jednadžbe stacionarne difuzije', under the supervision of Prof. Krešimir Burazin. In November 2013, she enrolled in the Croatian doctoral program in Mathematics at the Department of Mathematics, Faculty of Science at the University of Zagreb.

From May 2014, she has been employed as a teaching assistant at the Department of Mathematics, University of Osijek. She was one of the collaborators on two scientific projects of University of J. J. Strossmayer of Osijek: 'Evolution Friedrichs systems', under the leadership of Krešimir Burazin and 'Damping optimization in mechanics systems excited with external force', under the leadership of Zoran Tomljanović. She was also one of the collaborators on a bilateral project with Serbia, 'Calculus of variations, optimisation and application', cofinanced by the Serbian Ministry of Science, under the leadership of Nataša Krejić (University of Novi Sad) and Krešimir Burazin (University of Osijek). Currently, she is one of the collaborators on the scientific project 'Homogenization, dimension reduction and structural optimization in continuum mechanics', funded by the Croatian Science Foundation since 2018, with project leader Igor Velčić.

She attended more than 20 schools, workshops and conferences in Croatia and abroad, where she repeatedly gave talks. She is one of the authors of two published scientific papers and one accepted for publication.

List of publications

Journal Publications:

1. Antonić, A., Burazin, K., Crnjac, I. & Erceg, M. (2017), 'Complex Friedrichs systems and applications', *Journal of Mathematical Physics*, **58**(10), 1–22.

2. Burazin, K., Crnjac, I. & Vrdoljak, M. (2018), 'Variant of optimality criteria method for multiple state optimal design problems', *Communications in Mathematical Sciences*, **16**(6), 1597–1614.

3. Burazin, K. & Crnjac, I. (2019), ‘Convergence of the optimality criteria method for multiple state optimal design problems’ *Computers and Mathematics with Applications*, https://doi.org/10.1016/j.camwa.2019.09.002

Others:

1. Soldo, I. & Vuksanović, I. (2014), ‘Pitagorine trojke’, *Matematičko fizički list*, **255**, 179–184.

2. Burazin, K., Tomljanović, Z. & Vuksanović, I. (2014), ‘Prigušivanje mehaničkih vibracija’, *Math.e: hrvatski matematički elektronski časopis*, **24**