



# Optimization of the principal eigenvalue for elliptic operators

Hongwei Lou<sup>1</sup> · Jiongmin Yong<sup>2</sup>

Received: 29 November 2019 / Accepted: 31 March 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

Maximization and minimization problems of the principle eigenvalue for divergence form second order elliptic operators with the Dirichlet boundary condition are considered. The principal eigen map of such elliptic operators is introduced and some basic properties of this map, including continuity, concavity, and differentiability with respect to the parameter in the diffusibility matrix, are established. For maximization problem, the admissible control set is convexified to get the existence of an optimal convexified relaxed solution. Whereas, for minimization problem, the relaxation of the problem under  $H$ -convergence is introduced to get an optimal  $H$ -relaxed solution for certain interesting special cases. Some necessary optimality conditions are presented for both problems and a couple of illustrative examples are presented as well.

**Mathematics Subject Classification** 35J15 · 35P05 · 47A75 · 49K20 · 49J20

## 1 Introduction

Consider a heat conduct problem in a bounded domain  $\Omega \subseteq \mathbb{R}^n$ . Suppose  $\Omega$  is occupied by a certain type of medium with (not necessarily isotropic) uniformly elliptic *diffusibility matrix*  $a(\cdot) \equiv (a_{ij}(\cdot))$ . Let  $y(t, x)$  be the temperature of the body at  $(t, x)$ . Then, in the case that there is neither source nor sink of the heat in the domain, and the temperature is set to be a fixed level (say, 0, for simplicity) at the boundary  $\partial\Omega$ , the (temperature) function  $y(\cdot, \cdot) \equiv y(\cdot, \cdot; y_0(\cdot))$  will be the weak solution to the following parabolic equation:

---

Communicated by F. H. Lin.

---

This work is supported in part by NSFC Grant 11771097 and by NSF Grant DMS-1812921.

---

✉ Jiongmin Yong  
jiongmin.yong@ucf.edu

Hongwei Lou  
hwlou@fudan.edu.cn

<sup>1</sup> School of Mathematical Sciences, Fudan University, Shanghai 200433, China

<sup>2</sup> Department of Mathematics, University of Central Florida, Orlando, FL 32816, USA

$$\begin{cases} y_t(t, x) = \nabla \cdot (a(x) \nabla y(t, x)), & (t, x) \in [0, \infty) \times \Omega, \\ y(t, x) = 0, & (t, x) \in [0, \infty) \times \partial\Omega, \\ y(0, x) = y_0(x), & x \in \Omega. \end{cases} \quad (1.1)$$

Consequently,

$$\begin{aligned} \frac{d}{dt} \|y(t, \cdot)\|_2^2 &= 2 \int_{\Omega} y(t, x) \nabla \cdot (a(x) \nabla y(t, x)) dx \\ &= -2 \int_{\Omega} \langle a(x) \nabla y(t, x), \nabla y(t, x) \rangle dx \\ &\leq -2\lambda_{a(\cdot)} \int_{\Omega} |y(t, x)|^2 dx \equiv -2\lambda_{a(\cdot)} \|y(t, \cdot)\|_2^2, \end{aligned} \quad (1.2)$$

where  $\|\cdot\|_2 = \|\cdot\|_{L^2(\Omega)}$  is the usual  $L^2(\Omega)$ -norm, and  $\lambda_{a(\cdot)} > 0$  is the *smallest eigenvalue* (which is called the *principal eigenvalue*) of the differential operator  $\mathcal{L}_{a(\cdot)}$  defined by the following:

$$\begin{aligned} \mathcal{L}_{a(\cdot)}\varphi(\cdot) &= -\nabla \cdot (a(\cdot) \nabla \varphi(\cdot)), \quad \varphi(\cdot) \in \mathcal{D}(\mathcal{L}_{a(\cdot)}), \\ \mathcal{D}(\mathcal{L}_{a(\cdot)}) &= \left\{ \varphi(\cdot) \in W_0^{1,2}(\Omega) \mid \nabla \cdot (a(\cdot) \nabla \varphi(\cdot)) \in L^2(\Omega) \right\}. \end{aligned} \quad (1.3)$$

In the above,  $W_0^{1,2}(\Omega)$  is the usual Sobolev space which is the completion of  $C_0^\infty(\Omega)$  (smooth functions with compact supports in  $\Omega$ ) under the norm (see [1]):

$$\|\varphi(\cdot)\|_{W_0^{1,2}(\Omega)} = \left( \|\nabla \varphi(\cdot)\|_2^2 + \|\varphi(\cdot)\|_2^2 \right)^{\frac{1}{2}} < \infty.$$

It is known that with such a  $\lambda_{a(\cdot)} > 0$ , the following boundary value problem

$$\begin{cases} -\nabla \cdot (a(x) \nabla y_1(x)) = \lambda_{a(\cdot)} y_1(x), & \text{in } \Omega, \\ y_1|_{\partial\Omega} = 0 \end{cases}$$

admits a weak solution  $y_1(\cdot) \in W_0^{1,2}(\Omega) \setminus \{0\}$ , which is called a *principal eigenfunction* of operator  $\mathcal{L}_{a(\cdot)}$ . Moreover,

$$\lambda_{a(\cdot)} \equiv \inf_{y(\cdot) \in W_0^{1,2}(\Omega) \setminus \{0\}} \frac{\int_{\Omega} \langle a(x) \nabla y(x), \nabla y(x) \rangle dx}{\|y(\cdot)\|_2^2}. \quad (1.4)$$

From [15], Theorem 8.38, we know that the multiplicity of  $\lambda_{a(\cdot)}$  is 1, and  $y_1(\cdot) \equiv y_1(\cdot; a(\cdot))$  can be taken the unique eigenfunction such that it is positive in  $\Omega$  and normalized:

$$\|y_1(\cdot)\|_2^2 = \int_{\Omega} |y_1(x)|^2 dx = 1.$$

We call such a  $y_1(\cdot)$  the *normalized principal eigenfunction* of  $\mathcal{L}_{a(\cdot)}$  and denote it by  $y_{a(\cdot)}$ , indicating the dependence on  $a(\cdot)$ . For convenience, we call  $(\lambda_{a(\cdot)}, y_{a(\cdot)})$  the *normalized principal eigen-pair* of the operator  $\mathcal{L}_{a(\cdot)}$ . From (1.2), we see that with such a  $\lambda_{a(\cdot)}$ , one has

$$\|y(t, \cdot; y_0(\cdot))\|_2 \leq e^{-\lambda_{a(\cdot)} t} \|y_0(\cdot)\|_2, \quad t \geq 0, \quad \forall y_0(\cdot) \in L^2(\Omega)$$

and if the initial state  $y_0(\cdot) = y_{a(\cdot)}$ , then one has the equality:

$$y(t, x; y_{a(\cdot)}) = e^{-\lambda_{a(\cdot)} t} y_{a(\cdot)}(x), \quad (t, x) \in [0, \infty) \times \mathbb{R}^n,$$

which leads to

$$\|y(t, \cdot; y_{a(\cdot)})\|_2 = e^{-\lambda_{a(\cdot)} t} \|y_{a(\cdot)}\|_2 = e^{-\lambda_{a(\cdot)} t}, \quad t \geq 0.$$

Hence, one obtains the following representation:

$$\lambda_{a(\cdot)} = \inf_{y_0(\cdot) \in W_0^{1,2}(\Omega) \setminus \{0\}} \lim_{t \rightarrow \infty} \frac{1}{t} \log \frac{\|y_0(\cdot)\|_2}{\|y(t, \cdot; y_0(\cdot))\|_2} = -\frac{1}{t} \log \|y(t, \cdot; y_{a(\cdot)})\|_2, \quad \forall t > 0.$$

Consequently, in some sense,  $\lambda_{a(\cdot)}$  is the smallest (uniform) decay rate for the evolutionary map  $y_0(\cdot) \mapsto y(t, \cdot; y_0(\cdot))$  (uniform with respect to the initial state  $y_0(\cdot)$ ).

On the other hand, since  $a(\cdot)$  is assumed to be uniformly elliptic, the following Poincaré's inequality always holds:

$$\|y(\cdot)\|_2^2 \leq C \int_{\Omega} \langle a(x) \nabla y(x), \nabla y(x) \rangle dx, \quad \forall y(\cdot) \in W_0^{1,2}(\Omega) \quad (1.5)$$

for some constant  $C > 0$ . Thus, the sharp constant  $C$  that makes the above true is given by the following

$$\sup_{y(\cdot) \in W_0^{1,2}(\Omega) \setminus \{0\}} \frac{\|y(\cdot)\|_2^2}{\int_{\Omega} \langle a(x) \nabla y(x), \nabla y(x) \rangle dx} \equiv \frac{1}{\lambda_{a(\cdot)}} < \infty. \quad (1.6)$$

Therefore, the sharp Poincaré's inequality reads

$$\|y(\cdot)\|_2^2 \leq \frac{1}{\lambda_{a(\cdot)}} \int_{\Omega} \langle a(x) \nabla y(x), \nabla y(x) \rangle dx, \quad \forall y(\cdot) \in W_0^{1,2}(\Omega). \quad (1.7)$$

If the diffusibility matrix  $a(\cdot)$  can be chosen from a given set  $\mathcal{A}$ , which amounts to saying that the composite material/medium occupying  $\Omega$  can be designed within a certain range (c.f. [2, 11, 29, 34]), then we may try to minimize  $\lambda_{a(\cdot)}$  (preserving the temperature of the body in a certain fashion), or to maximize  $\lambda_{a(\cdot)}$  (cooling down the body as quick as possible, uniformly in the initial temperature distribution). In terms of Poincaré's inequality, the former means that we are looking for the sharp constant uniform for  $a(\cdot) \in \mathcal{A}$ , and the latter means that we are looking for the smallest possible sharp constant for some  $a(\cdot) \in \mathcal{A}$ .

Now, let  $0 < \mu_0 \leq \mu_1 < \infty$  be given and let

$$M[\mu_0, \mu_1] = \left\{ A \in \mathbb{S}^n \mid \mu_0 I \leq A \leq \mu_1 I \right\}, \quad (1.8)$$

where  $\mathbb{S}^n$  is the set of all  $(n \times n)$  symmetric matrices. Define

$$\mathcal{M}[\mu_0, \mu_1] = \left\{ a : \Omega \rightarrow M[\mu_0, \mu_1] \mid a(\cdot) \text{ is measurable} \right\} \subseteq L^\infty(\Omega; \mathbb{S}^n). \quad (1.9)$$

Clearly,  $M[\mu_0, \mu_1]$  is convex and compact in  $\mathbb{S}^n$ . Consequently,  $\mathcal{M}[\mu_0, \mu_1]$  is convex and closed in  $L^p(\Omega; \mathbb{S}^n)$  (for any  $p \in [1, \infty]$ ). From (1.4), we see that

$$\lambda_{a(\cdot)} \leq \lambda_{\bar{a}(\cdot)}, \quad \forall a(\cdot), \bar{a}(\cdot) \in \mathcal{M}[\mu_0, \mu_1], \quad a(\cdot) \leq \bar{a}(\cdot), \quad (1.10)$$

namely, the map  $a(\cdot) \mapsto \lambda_{a(\cdot)}$  is monotone non-decreasing. In particular,

$$\mu_0 \lambda_I \leq \lambda_{a(\cdot)} \leq \mu_1 \lambda_I, \quad \forall a(\cdot) \in \mathcal{M}[\mu_0, \mu_1], \quad (1.11)$$

where  $\lambda_I$  is the principal eigenvalue of  $-\Delta$  on  $\Omega$ , with the homogeneous Dirichlet boundary condition. Consequently,

$$\sup_{a(\cdot) \in \mathcal{M}[\mu_0, \mu_1]} \lambda_{a(\cdot)} = \mu_1 \lambda_I, \quad \inf_{a(\cdot) \in \mathcal{M}[\mu_0, \mu_1]} \lambda_{a(\cdot)} = \mu_0 \lambda_I. \quad (1.12)$$

This shows that minimizing or maximizing  $\lambda_{a(\cdot)}$  over  $\mathcal{M}[\mu_0, \mu_1]$  is trivial. Now, we let  $\mathcal{A}$  be chosen with

$$\emptyset \neq \mathcal{A} \subseteq \mathcal{M}[\mu_0, \mu_1], \quad 0 < \mu_0 \leq \mu_1 < \infty, \quad (1.13)$$

and pose the following two problems.

**Problem** ( $\bar{\Lambda}(\mathcal{A})$ ). Find an  $\bar{a}(\cdot) \in \mathcal{A}$  such that

$$\lambda_{\bar{a}(\cdot)} = \sup_{a(\cdot) \in \mathcal{A}} \lambda_{a(\cdot)}. \quad (1.14)$$

**Problem** Problem ( $\underline{\Lambda}(\mathcal{A})$ ). Find an  $\underline{a}(\cdot) \in \mathcal{A}$  such that

$$\lambda_{\underline{a}(\cdot)} = \inf_{a(\cdot) \in \mathcal{A}} \lambda_{a(\cdot)}. \quad (1.15)$$

Some general results will be presented concerning the above two problems in the next section.

Further, to obtain finer results, we will concentrate on a more specific case which we now describe. Fix two different matrices  $A_0, A_1 \in M[\mu_0, \mu_1]$  with  $0 < \mu_0 \leq \mu_1 < \infty$ . We define

$$A(r) = (1 - r)A_0 + rA_1, \quad r \in [0, 1], \quad (1.16)$$

and for some  $0 \leq \alpha \leq \beta \leq 1$ , let

$$\mathcal{U}[\alpha, \beta] = \left\{ \chi_{\Omega_1}(\cdot) \mid \Omega_1 \subseteq \Omega \text{ is measurable, } \alpha|\Omega| \leq |\Omega_1| \leq \beta|\Omega| \right\}, \quad (1.17)$$

where  $\chi_{\Omega_1}(\cdot)$  is the characteristic function of  $\Omega_1$  and  $|\Omega_1|$  is the Lebesgue measure of measurable set  $\Omega_1$ . Then for any  $u(\cdot) \equiv \chi_{\Omega_1}(\cdot) \in \mathcal{U}[\alpha, \beta]$ , we have

$$A(u(\cdot)) = [1 - \chi_{\Omega_1}(\cdot)]A_0 + \chi_{\Omega_1}(\cdot)A_1 = A_0 + \chi_{\Omega_1}(\cdot)(A_1 - A_0).$$

Denote

$$\mathcal{A}[\alpha, \beta] = A(\mathcal{U}[\alpha, \beta]) = \left\{ A_0 + \chi_{\Omega_1}(\cdot)(A_1 - A_0) \mid \Omega_1 \text{ measurable, } \alpha|\Omega| \leq |\Omega_1| \leq \beta|\Omega| \right\}. \quad (1.18)$$

Note that the set  $\mathcal{A}[\alpha, \beta]$  is non-convex (unless  $A_0 = A_1$  which is excluded). For the heat conduct problem, with  $a(\cdot) = A(u(\cdot))$  for some  $u(\cdot) = \chi_{\Omega_1}(\cdot) \in \mathcal{U}[\alpha, \beta]$ , it means that two media occupy the domain, the one with conductivity matrix  $A_1$  occupies  $\Omega_1$  and the other with conductivity matrix  $A_0$  occupies  $\Omega \setminus \Omega_1$ . The corresponding principal eigenvalue and the corresponding (unique) normalized principal eigenfunction are denoted by

$$\lambda_{u(\cdot)} \equiv \lambda_{A_0 + u(\cdot)(A_1 - A_0)}, \quad y_{u(\cdot)} \equiv y_{A_0 + u(\cdot)(A_1 - A_0)},$$

and the following holds:

$$\begin{cases} -\nabla \cdot ( [A_0 + u(x)(A_1 - A_0)] \nabla y_{u(\cdot)}(x) ) = \lambda_{u(\cdot)} y_{u(\cdot)}(x), & x \in \Omega, \\ y_{u(\cdot)}|_{\partial\Omega} = 0. \end{cases} \quad (1.19)$$

Then we can pose the following problems.

**Problem** ( $\bar{\Lambda}[\alpha, \beta]$ ). For given  $0 \leq \alpha \leq \beta \leq 1$ , find a  $\bar{u}(\cdot) \in \mathcal{U}[\alpha, \beta]$  such that

$$\lambda_{\bar{u}(\cdot)} = \sup_{u(\cdot) \in \mathcal{U}[\alpha, \beta]} \lambda_{u(\cdot)}. \quad (1.20)$$

**Problem** ( $\underline{\Lambda}[\alpha, \beta]$ ). For given  $0 \leq \alpha \leq \beta \leq 1$ , find a  $\underline{u}(\cdot) \in \mathcal{U}[\alpha, \beta]$  such that

$$\lambda_{\underline{u}(\cdot)} = \inf_{u(\cdot) \in \mathcal{U}[\alpha, \beta]} \lambda_{u(\cdot)}. \quad (1.21)$$

Any  $\bar{u}(\cdot) \in \mathcal{U}[\alpha, \beta]$  (resp.  $\underline{u}(\cdot) \in \mathcal{U}[\alpha, \beta]$ ) satisfying (1.20) (resp. (1.21)) is called an *optimal control* of Problem ( $\bar{\Lambda}[\alpha, \beta]$ ) (resp. Problem ( $\underline{\Lambda}[\alpha, \beta]$ )).

Note that if  $0 = \alpha < \beta \leq 1$ , and, say,  $A_0 \leq A_1$ , due to a fact similar to (1.11)–(1.12), one has

$$\inf_{u(\cdot) \in \mathcal{U}[0, \beta]} \lambda(u(\cdot)) = \lambda_{A_0}, \quad (1.22)$$

making Problem ( $\underline{\Lambda}[0, \beta]$ ) trivial; and likewise, if  $0 < \alpha \leq \beta = 1$ , and still let  $A_0 \leq A_1$ , then

$$\sup_{u(\cdot) \in \mathcal{U}[\alpha, 1]} \lambda(u(\cdot)) = \lambda_{A_1}, \quad (1.23)$$

making Problem ( $\bar{\Lambda}[\alpha, 1]$ ) trivial. To avoid such situations, in what follows, we will assume the following:

$$\begin{cases} \text{either } 0 < \alpha \leq \beta < 1, \text{ no additional restrictions on } A_0, A_1, \\ \text{or } \alpha = 0, \beta = 1, \text{ and neither } A_0 \leq A_1, \text{ nor } A_1 \leq A_0 \text{ holds.} \end{cases} \quad (1.24)$$

Note that when  $0 < \alpha \leq \beta < 1$ , even if, say,  $A_1 = 2A_0 > A_0$ , the location/shape of the optimal  $\Omega_1$  (if it exists, which is non-empty and not equal to  $\Omega$ ) is not obvious. On the other hand, in the case that  $A_0$  and  $A_1$  are not comparable, one expects that neither  $u_0(\cdot) = 0$  nor  $u_1(\cdot) = 1$  is optimal.

For either case in (1.24),  $\mathcal{U}[\alpha, \beta]$  is not convex. Hence, the existence of optimal controls for Problems ( $\bar{\Lambda}[\alpha, \beta]$ ) and ( $\underline{\Lambda}[\alpha, \beta]$ ) is not guaranteed, in general. To study these problems, we will introduce suitable relaxed problems for which the relaxed optimal controls will exist. Some necessary conditions for relaxed optimal controls will then be established, and illustrative examples will be presented as well.

It is possible to discuss the situation of more than two materials instead of just two. The main idea will be the same and the abstract results will have similar looking. However, further analysis on the optimal controls will be much more technical and complicated. Therefore, for the sake of simplification in our presentation, we prefer not to pursue that.

Some studies on optimization of the principal eigenvalue for elliptic operators can be found in the book by Henrot [18] (see also the references cited therein). The case studied in [18] was isotropic, namely, the diffusion matrix  $a(x) = \sigma(x)I$ , for some scalar function  $\sigma(\cdot)$ . Moreover, even for that case, only a maximization problem was considered. For other relevant works, here is a partial list of references: [5, 6, 8, 10, 12–14, 27].

The rest of this paper is organized as follows. Section 2 will be devoted to some general considerations of the problems that we are interested in. In Sect. 3, a convexification of maximization problem is investigated. In Sect. 4, a relaxation of minimization problem in terms of the so-called  $H$ -convergence will be studied. A detailed example is worked out in Sect. 5. Finally some remarks are collected in Sect. 6.

## 2 The principal Eigen map and its properties

We fix a bounded Lipschitz domain  $\Omega \subset \mathbb{R}^n$ , i.e.,  $\Omega \subseteq \mathbb{R}^n$  is a bounded domain with a Lipschitz boundary  $\partial\Omega$ , and constants  $0 < \mu_0 \leq \mu_1 < \infty$ . For any  $a(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ ,

recall that  $\mathcal{L}_{a(\cdot)}$  is an elliptic operator defined by (1.3), and  $(\lambda_{a(\cdot)}, y_{a(\cdot)})$  is the normalized principal eigen-pair of  $\mathcal{L}_{a(\cdot)}$ . Thus,

$$\begin{cases} -\nabla \cdot (a(x) \nabla y_{a(\cdot)}(x)) = \lambda_{a(\cdot)} y_{a(\cdot)}(x), & \text{in } \Omega, \\ y_{a(\cdot)}|_{\partial\Omega} = 0, \end{cases} \quad (2.1)$$

with  $y_{a(\cdot)}(x) > 0$  for  $x \in \Omega$ , and  $\|y_{a(\cdot)}(\cdot)\|_2 = 1$ . Further, it is known that

$$\lambda_{a(\cdot)} = \min_{y(\cdot) \in W_0^{1,2}(\Omega), \|y(\cdot)\|_2=1} \int_{\Omega} \langle a(x) \nabla y(x), \nabla y(x) \rangle dx = \int_{\Omega} \langle a(x) \nabla y_{a(\cdot)}(x), \nabla y_{a(\cdot)}(x) \rangle dx. \quad (2.2)$$

Define  $\Lambda : \mathcal{M}[\mu_0, \mu_1] \rightarrow \mathbb{R} \times W_0^{1,2}(\Omega)$  by

$$\Lambda(a(\cdot)) = (\lambda_{a(\cdot)}, y_{a(\cdot)}), \quad \forall a(\cdot) \in \mathcal{M}[\mu_0, \mu_1] \quad (2.3)$$

and call it the *principal eigen map* of the operator  $\mathcal{L}_{a(\cdot)}$ . We first recall the following result found in [19].

**Proposition 2.1** *Let  $a(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ . Then there exists a constant  $\gamma = \gamma(\mu_0, \mu_1; \Omega) > 0$  such that*

$$\lambda - \lambda_{a(\cdot)} \geq \gamma, \quad (2.4)$$

where  $\lambda$  is any eigenvalue of  $\mathcal{L}_{a(\cdot)}$  different from  $\lambda_{a(\cdot)}$ . Consequently, one has:

$$\int_{\Omega} \langle a(x) \nabla y(x), \nabla y(x) \rangle dx \geq (\lambda_{a(\cdot)} + \gamma) \|y(\cdot)\|_2^2, \quad \forall y(\cdot) \in W_0^{1,2}(\Omega), \quad \int_{\Omega} y(x) y_{a(\cdot)}(x) dx = 0. \quad (2.5)$$

Next, we present the following simple result which will be useful below.

**Proposition 2.2** *The map  $a(\cdot) \mapsto \lambda_{a(\cdot)}$  is concave on  $\mathcal{M}[\mu_0, \mu_1]$ . Consequently, this map is Lipschitz on  $L^\infty(\Omega; \mathbb{S}^n)$ , i.e., for some  $L > 0$ ,*

$$|\lambda_{a(\cdot)} - \lambda_{\bar{a}(\cdot)}| \leq L \|a(\cdot) - \bar{a}(\cdot)\|_\infty, \quad \forall a(\cdot), \bar{a}(\cdot) \in L^\infty(\Omega; \mathbb{S}^n). \quad (2.6)$$

**Proof** For any  $a_0(\cdot), a_1(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$  and any  $\gamma \in (0, 1)$ ,

$$\begin{aligned} & \lambda_{(1-\gamma)a_0(\cdot) + \gamma a_1(\cdot)} \\ &= \min_{\substack{y(\cdot) \in W_0^{1,2}(\Omega) \\ \|y(\cdot)\|_2=1}} \int_{\Omega} \langle [(1-\gamma)a_0(x) + \gamma a_1(x)] \nabla y(x), \nabla y(x) \rangle dx \\ &\geq (1-\gamma) \min_{\substack{y(\cdot) \in W_0^{1,2}(\Omega) \\ \|y(\cdot)\|_2=1}} \int_{\Omega} \langle a_0(x) \nabla y(x), \nabla y(x) \rangle dx \\ &\quad + \gamma \min_{\substack{y(\cdot) \in W_0^{1,2}(\Omega) \\ \|y(\cdot)\|_2=1}} \int_{\Omega} \langle a_1(x) \nabla y(x), \nabla y(x) \rangle dx \\ &\geq (1-\gamma)\lambda_{a_0(\cdot)} + \gamma\lambda_{a_1(\cdot)}. \end{aligned}$$

This proves the concavity of the map  $a(\cdot) \mapsto \lambda_{a(\cdot)}$ . Then the Lipschitz continuity follows from a standard argument (see, for example, [22], p.235, Lemma 2.8).  $\square$

We will see that the map  $a(\cdot) \mapsto \lambda_{a(\cdot)}$  is not strictly convex.

The following theorem is due to Gallouet–Monier [16], which is an extension of a result by Meyers [26].

**Theorem 2.3** Let  $\Omega \subseteq \mathbb{R}^n$  be a bounded Lipschitz domain and  $0 < \mu_0 \leq \mu_1$ . Then there exists a  $p_0 > 2$ , only depending on  $\Omega$ , such that for any  $2 < p < p_0$ ,  $a(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ ,  $\mathbf{f}(\cdot) \in L^p(\Omega)$  and  $h(\cdot) \in L^{\frac{np}{n+p}}(\Omega)$ , the following problem:

$$\begin{cases} -\nabla \cdot (a(x) \nabla y(x)) = \nabla \cdot \mathbf{f}(x) + h(x), & x \in \Omega, \\ y|_{\partial\Omega} = 0 \end{cases}$$

admits a unique weak solution  $y(\cdot) \in W_0^{1,p}(\Omega)$ , and the following estimate holds:

$$\|\nabla y(\cdot)\|_p \leq C \left( \|\mathbf{f}(\cdot)\|_p + \|h(\cdot)\|_{\frac{np}{n+p}} \right). \quad (2.7)$$

Hereafter,  $C > 0$  represents a generic constant which could be different from line to line.

Note that in the above result,  $p > 2$ , which will play a crucial role below. The following result is concerned with the principal eigen map  $\Lambda$ .

**Theorem 2.4** Let  $0 < \mu_0 \leq \mu_1 < \infty$  and  $\Omega \subseteq \mathbb{R}^n$  be a Lipschitz domain. Then there exists a  $p > 2$  such that the principal eigen map  $\Lambda$  is Lipschitz continuous in the following sense:

$$|\lambda_{a_1(\cdot)} - \lambda_{a_2(\cdot)}| \leq C \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p}{p-2}}, \quad \forall a_1(\cdot), a_2(\cdot) \in \mathcal{M}[\mu_0, \mu_1], \quad (2.8)$$

and for any  $\bar{p} \in [2, p)$ ,

$$\|y_{a_1(\cdot)} - y_{a_2(\cdot)}\|_{W^{1,\bar{p}}(\Omega)} \leq C \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}}, \quad \forall a_1(\cdot), a_2(\cdot) \in \mathcal{M}[\mu_0, \mu_1]. \quad (2.9)$$

Note that (2.8) is an improvement of (2.6), thanks to the existence of a  $p > 2$  so that (2.7) holds.

**Proof** First of all, for any  $a(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ , we recall that

$$\mu_0 \|\nabla y_{a(\cdot)}\|_2^2 \leq \int_{\Omega} \langle a(x) \nabla y_{a(\cdot)}(x), \nabla y_{a(\cdot)}(x) \rangle dx = \lambda_{a(\cdot)} \leq \mu_1 \lambda_I$$

and by Sobolev embedding theorem ([1]), for any  $n \geq 2$ , we have

$$W^{1,2}(\Omega) \hookrightarrow L^r(\Omega), \quad r < \frac{2n}{n-2}$$

with the convention that  $\frac{2n}{n-2} = \infty$  when  $n = 2$ .

Now, let  $p_0 > 0$  be the number in Theorem 2.3 and chose  $2 < p < \min(p_0, \frac{2n}{n-2})$ . We have  $p(n-2) < 2n$ , which leads to  $\frac{np}{n+p} < 2$ . Hence, by Theorem 2.3, regarding  $\lambda_{a(\cdot)} y_{a(\cdot)}$  as a nonhomogeneous term on the right-hand side of the equation (2.1), we have (noting  $\|y_{a(\cdot)}\|_2 = 1$ )

$$\|y_{a(\cdot)}\|_{W^{1,p}(\Omega)} \leq C_p \lambda_{a(\cdot)} \|y_{a(\cdot)}\|_{\frac{np}{n+p}} \leq C_p \lambda_{a(\cdot)} |\Omega|^{\frac{1}{n} + \frac{1}{p} - \frac{1}{2}} = \bar{C}_p. \quad (2.10)$$

Here  $\bar{C}_p$  is an absolute constant, uniform in  $a(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ . Let

$$\mathcal{Y}^p(\mu_0, \mu_1) = \{y(\cdot) \in W_0^{1,p}(\Omega) \mid \|y(\cdot)\|_2 = 1, \|y(\cdot)\|_{W^{1,p}(\Omega)} \leq \bar{C}_p, y(x) \geq 0, \text{ a.e.}\}.$$

Then  $y_{a(\cdot)} \in \mathcal{Y}^p(\mu_0, \mu_1) \subseteq W_0^{1,2}(\Omega)$ . Consequently, for any  $a_1(\cdot), a_2(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ , denote

$$\lambda_i = \lambda_{a_i(\cdot)}, \quad y_i(\cdot) = y_{a_i(\cdot)}(\cdot), \quad i = 1, 2,$$

and assume that  $\lambda_1 \geq \lambda_2$ . We have

$$\begin{aligned}\lambda_1 - \lambda_2 &= \lambda_1 - \int_{\Omega} \langle a_2(x) \nabla y_2(x), \nabla y_2(x) \rangle dx \\ &\leq \int_{\Omega} \langle a_1(x) \nabla y_2(x), \nabla y_2(x) \rangle dx - \int_{\Omega} \langle a_2(x) \nabla y_2(x), \nabla y_2(x) \rangle dx \\ &\leq \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p}{p-2}} \|\nabla y_2(\cdot)\|_p^2 \leq \tilde{C}_p^2 \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p}{p-2}},\end{aligned}$$

proving (2.8). Next, let

$$\begin{aligned}\alpha &= \int_{\Omega} y_1(\xi) y_2(\xi) d\xi, \\ y_{12}(x) &= y_1(x) - \alpha y_2(x) \quad y_{21}(x) = y_2(x) - \alpha y_1(x).\end{aligned}$$

Then  $\alpha \geq 0$ , and

$$\begin{aligned}\int_{\Omega} y_{12}(x) y_2(x) dx &= \int_{\Omega} y_{21}(x) y_1(x) dx = 0, \quad \|y_{12}(\cdot)\|_2^2 = \|y_{21}(\cdot)\|_2^2 = 1 - \alpha^2, \\ (1 + \alpha)(y_2(x) - y_1(x)) &= y_{21}(x) - y_{12}(x).\end{aligned}$$

Hence,

$$\|y_2(\cdot) - y_1(\cdot)\|_2^2 \leq \|y_{21}(\cdot) - y_{12}(\cdot)\|^2 \leq 2\|y_{21}(\cdot)\|_2^2 + 2\|y_{12}(\cdot)\|^2 = 4(1 - \alpha^2). \quad (2.11)$$

By (2.5) and noting that

$$\int_{\Omega} \langle (a_1(x) - a_2(x)) \nabla y_1(x), \nabla y_2(x) \rangle = \lambda_1 \int_{\Omega} y_1 y_2 - \lambda_2 \int_{\Omega} y_1 y_2 = (\lambda_1 - \lambda_2) \alpha,$$

one has

$$\begin{aligned}(\lambda_1 + \gamma)(1 - \alpha^2) &\leq \int_{\Omega} \langle a_1(x) \nabla y_{21}(x), \nabla y_{21}(x) \rangle dx \\ &= \int_{\Omega} \langle a_1(x) \nabla y_2(x), \nabla y_2(x) \rangle dx - 2\alpha \int_{\Omega} \langle a_1(x) \nabla y_1(x), \nabla y_2(x) \rangle dx \\ &\quad + \alpha^2 \int_{\Omega} \langle a_1(x) \nabla y_1(x), \nabla y_1(x) \rangle dx \\ &= \int_{\Omega} \langle (a_1(x) - a_2(x)) \nabla y_2(x), \nabla y_2(x) \rangle dx \\ &\quad + \lambda_2 - 2\lambda_1 \alpha \int_{\Omega} y_1(x) y_2(x) dx + \alpha^2 \lambda_1 \\ &= \int_{\Omega} \langle (a_1(x) - a_2(x)) \nabla (y_2(x) - y_1(x)), \nabla y_2(x) \rangle dx + (\lambda_1 - \lambda_2) \alpha + \lambda_2 - \alpha^2 \lambda_1 \\ &\leq \|a_1(\cdot) - a_2(\cdot)\|_{\frac{2p}{p-2}} \|\nabla y_1(\cdot) - \nabla y_2(\cdot)\|_2 \|\nabla y_2(\cdot)\|_p \\ &\quad + (1 - \alpha)(\lambda_2 - \lambda_1) + (1 - \alpha^2) \lambda_1.\end{aligned}$$

Consequently, since  $\alpha \leq 1$  and it is assumed that  $\lambda_2 \leq \lambda_1$ ,

$$(1 - \alpha^2) \gamma \leq C \|a_1(\cdot) - a_2(\cdot)\|_{\frac{2p}{p-2}} \|\nabla y_1(\cdot) - \nabla y_2(\cdot)\|_2.$$

Combining the above with (2.11), we have

$$\|y_1(\cdot) - y_2(\cdot)\|_2^2 \leq C \|a_1(\cdot) - a_2(\cdot)\|_{\frac{2p}{p-2}} \|\nabla y_1(\cdot) - \nabla y_2(\cdot)\|_2. \quad (2.12)$$

On the other hand,

$$\begin{cases} -\nabla \cdot (a_1(x) \nabla [y_1(x) - y_2(x)]) = \lambda_1 y_1 - \lambda_2 y_2 + \nabla \cdot ([a_1(x) - a_2(x)] \nabla y_2(x)), \\ (y_1 - y_2)|_{\partial\Omega} = 0. \end{cases}$$

It follows from Theorem 2.3 that (with  $2 \leq \bar{p} < p$ )

$$\begin{aligned} & \|y_1(\cdot) - y_2(\cdot)\|_{W^{1,\bar{p}}(\Omega)} \\ & \leq C \| (a_1(\cdot) - a_2(\cdot)) \nabla y_2 \|_{\bar{p}} + C \|\lambda_1 y_1(\cdot) - \lambda_2 y_2(\cdot)\|_{\frac{n\bar{p}}{n+\bar{p}}} \\ & \leq C \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}} \|\nabla y_2(\cdot)\|_p + C |\Omega|^{\frac{1}{n} + \frac{1}{\bar{p}} - \frac{1}{2}} \|\lambda_1 y_1(\cdot) - \lambda_2 y_2(\cdot)\|_2 \\ & \leq C \bar{C}_p \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}} + C |\Omega|^{\frac{1}{n} + \frac{1}{\bar{p}} - \frac{1}{2}} (|\lambda_1 - \lambda_2| + \mu_1 \|y_1(\cdot) - y_2(\cdot)\|_2) \\ & \leq C \left( \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}} + \|y_1(\cdot) - y_2(\cdot)\|_2 \right). \end{aligned} \quad (2.13)$$

Applying (2.12), we have

$$\begin{aligned} & \|y_1(\cdot) - y_2(\cdot)\|_{W^{1,\bar{p}}(\Omega)} \\ & \leq C \left( \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}} + \|a_1(\cdot) - a_2(\cdot)\|_{\frac{2p}{p-2}}^{\frac{1}{2}} \|y_1(\cdot) - y_2(\cdot)\|_{W^{1,2}(\Omega)}^{\frac{1}{2}} \right) \\ & \leq C \left( \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}} + \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}}^{\frac{1}{2}} \|y_1(\cdot) - y_2(\cdot)\|_{W^{1,\bar{p}}(\Omega)}^{\frac{1}{2}} \right) \\ & \leq C \|a_1(\cdot) - a_2(\cdot)\|_{\frac{p\bar{p}}{p-\bar{p}}} + \frac{1}{2} \|y_1(\cdot) - y_2(\cdot)\|_{W^{1,\bar{p}}(\Omega)}. \end{aligned}$$

Then (2.9) follows.  $\square$

We now look at the directional differentiability of the principal eigen map  $\Lambda$ .

**Proposition 2.5** *Let  $\Omega$  be a bounded Lipschitz domain and  $a(\cdot), \bar{a}(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ . Then the directional derivative of the eigen map  $\Lambda$  at  $\bar{a}(\cdot)$  in the direction of  $a(\cdot) - \bar{a}(\cdot)$  is given by*

$$\Lambda'(\bar{a}(\cdot); a(\cdot) - \bar{a}(\cdot)) \equiv \lim_{\varepsilon \downarrow 0} \frac{\Lambda(\bar{a}(\cdot) + \varepsilon[a(\cdot) - \bar{a}(\cdot)]) - \Lambda(\bar{a}(\cdot))}{\varepsilon} = (\lambda', y'(\cdot)), \quad \text{in } \mathbb{R} \times W_0^{1,2}(\Omega), \quad (2.14)$$

where  $\lambda'$  is given by the following:

$$\lambda' = \int_{\Omega} \langle [a(x) - \bar{a}(x)] \nabla y_{\bar{a}(\cdot)}(x), \nabla y_{\bar{a}(\cdot)}(x) \rangle dx, \quad (2.15)$$

and  $y'(\cdot)$  is the unique weak solution of

$$\begin{cases} -\nabla \cdot (\bar{a}(x) \nabla y'(x)) = \lambda_{\bar{a}(\cdot)} y'(x) + \lambda' y_{\bar{a}(\cdot)}(x) + \nabla \cdot ([a(x) - \bar{a}(x)] \nabla y_{\bar{a}(\cdot)}(x)), \\ y'|_{\partial\Omega} = 0 \end{cases} \quad (2.16)$$

with the following additional condition:

$$\int_{\Omega} y'(x) y_{\bar{a}(\cdot)}(x) dx = 0. \quad (2.17)$$

**Proof** Let  $\varepsilon \in (0, 1)$ . Denote

$$a_\varepsilon(\cdot) = \bar{a}(\cdot) + \varepsilon[a(\cdot) - \bar{a}(\cdot)] \in \mathcal{M}(\mu_0, \mu_1),$$

and

$$\Lambda(\bar{a}(\cdot)) = (\lambda_{\bar{a}(\cdot)}, y_{\bar{a}(\cdot)}) \equiv (\bar{\lambda}, \bar{y}(\cdot)), \quad \Lambda(a_\varepsilon(\cdot)) = (\lambda_{a_\varepsilon(\cdot)}, y_{a_\varepsilon(\cdot)}) \equiv (\lambda_\varepsilon, y_\varepsilon(\cdot)).$$

Let

$$\lambda'_\varepsilon = \frac{\lambda_\varepsilon - \bar{\lambda}}{\varepsilon}, \quad y'_\varepsilon(\cdot) = \frac{y_\varepsilon(\cdot) - \bar{y}(\cdot)}{\varepsilon}.$$

Note that

$$\begin{aligned} \frac{a_\varepsilon(x)\nabla y_\varepsilon(x) - \bar{a}(x)\nabla \bar{y}(x)}{\varepsilon} &= \frac{[\bar{a}(x) + \varepsilon(a(x) - \bar{a}(x))]\nabla y_\varepsilon(x) - \bar{a}(x)\nabla \bar{y}(x)}{\varepsilon} \\ &= \bar{a}(x)\nabla y'_\varepsilon(x) + [a(x) - \bar{a}(x)]\nabla y_\varepsilon(x), \end{aligned}$$

and

$$\frac{\lambda_\varepsilon y_\varepsilon(x) - \bar{\lambda} \bar{y}(x)}{\varepsilon} = \frac{\lambda_\varepsilon [y_\varepsilon(x) - \bar{y}(x)] + (\lambda_\varepsilon - \bar{\lambda}) \bar{y}(x)}{\varepsilon} = \lambda_\varepsilon y'_\varepsilon(x) + \lambda'_\varepsilon \bar{y}(x).$$

Hence,  $(\lambda'_\varepsilon, y'_\varepsilon(\cdot))$  satisfies

$$\begin{cases} -\nabla \cdot (\bar{a}(x)\nabla y'_\varepsilon(x)) = \lambda_\varepsilon y'_\varepsilon(x) + \lambda'_\varepsilon \bar{y}(x) + \nabla \cdot ([a(x) - \bar{a}(x)]\nabla y_\varepsilon(x)), \\ y'_\varepsilon|_{\partial\Omega} = 0. \end{cases} \quad (2.18)$$

Since  $a(\cdot) \mapsto \lambda_{a(\cdot)}$  is Lipschitz, we have

$$|\lambda'_\varepsilon| = \frac{|\lambda_\varepsilon - \bar{\lambda}|}{\varepsilon} \leq L \|a(\cdot) - \bar{a}(\cdot)\|_\infty.$$

From the Lipschitz continuity of the principal eigen map  $\Lambda$  (see Theorem 2.4), we have

$$|\lambda'_\varepsilon| + \|y'_\varepsilon(\cdot)\|_{W_0^{1,2}(\Omega)} \leq C, \quad \forall \varepsilon > 0. \quad (2.19)$$

Then along a sequence  $\varepsilon_k \rightarrow 0^+$ , it holds that

$$\lambda'_{\varepsilon_k} \rightarrow \lambda', \quad \text{and} \quad y'_{\varepsilon_k}(\cdot) \rightarrow y'(\cdot), \quad \text{weakly in } W_0^{1,2}(\Omega).$$

Also, (2.19) leads to

$$\lambda_{\varepsilon_k} \rightarrow \bar{\lambda}, \quad \text{and} \quad y_{\varepsilon_k}(\cdot) \rightarrow \bar{y}(\cdot), \quad \text{strongly in } W_0^{1,2}(\Omega).$$

Consequently, passing to the limit in (2.18), one gets

$$\begin{cases} -\nabla \cdot (\bar{a}(x)\nabla y'(x)) = \bar{\lambda} y'(x) + \lambda' \bar{y}(x) + \nabla \cdot ([a(x) - \bar{a}(x)]\nabla \bar{y}(x)), \\ y'|_{\partial\Omega} = 0, \end{cases} \quad (2.20)$$

and

$$\begin{aligned} \lambda' &= \lambda' \|\bar{y}(\cdot)\|_2^2 = \int_{\Omega} \bar{y}(x) \left( -\nabla \cdot (\bar{a}(x)\nabla y'(x)) - \bar{\lambda} y'(x) - \nabla \cdot ([a(x) - \bar{a}(x)]\nabla \bar{y}(x)) \right) dx \\ &= \int_{\Omega} \left( \langle \bar{a}(x)\nabla \bar{y}(x), \nabla y'(x) \rangle - \bar{\lambda} \bar{y}(x) y'(x) + \langle \nabla \bar{y}(x), [a(x) - \bar{a}(x)]\nabla \bar{y}(x) \rangle \right) dx \\ &= \int_{\Omega} \langle [a(x) - \bar{a}(x)]\nabla \bar{y}(x), \nabla \bar{y}(x) \rangle dx. \end{aligned}$$

This implies that  $\lambda'_\varepsilon \rightarrow \lambda'$  for the whole sequence  $\varepsilon \rightarrow 0^+$ .

On the other hand, we have

$$\begin{aligned} 0 &\leq \lim_{k \rightarrow \infty} \frac{1}{\varepsilon_k} \int_{\Omega} \left( \langle \bar{a}(x) \nabla y_{\varepsilon_k}(x), \nabla y_{\varepsilon_k}(x) \rangle - \langle \bar{a}(x) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \right) dx \\ &= \lim_{k \rightarrow \infty} \int_{\Omega} \langle \bar{a}(x) \nabla y'_{\varepsilon_k}(x), \nabla y_{\varepsilon_k}(x) + \nabla \bar{y}(x) \rangle dx \\ &= 2 \int_{\Omega} \langle \bar{a}(x) \nabla y'(x), \nabla \bar{y}(x) \rangle dx = 2\bar{\lambda} \int_{\Omega} y'(x) \bar{y}(x) dx. \end{aligned}$$

Similarly,

$$\begin{aligned} 0 &\geq \lim_{k \rightarrow \infty} \frac{1}{\varepsilon_k} \int_{\Omega} \left( \langle a_{\varepsilon_k}(x) \nabla y_{\varepsilon_k}(x), \nabla y_{\varepsilon_k}(x) \rangle - \langle a_{\varepsilon_k}(x) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \right) dx \\ &= \lim_{k \rightarrow \infty} \int_{\Omega} \langle a_{\varepsilon_k}(x) \nabla y'_{\varepsilon_k}(x), \nabla y_{\varepsilon_k}(x) + \nabla \bar{y}(x) \rangle dx \\ &= 2 \int_{\Omega} \langle \bar{a}(x) \nabla y'(x), \nabla \bar{y}(x) \rangle dx = 2\bar{\lambda} \int_{\Omega} y'(x) \bar{y}(x) dx. \end{aligned}$$

Therefore, (2.17) holds. If  $\tilde{y}'(\cdot)$  satisfying (2.17) is also a solution of (2.16), then  $\tilde{y}'(\cdot) - y'(\cdot)$  is either zero or a principal eigenfunction of  $\mathcal{L}_{\bar{a}(\cdot)}$ . Thus  $\tilde{y}'(\cdot) - y'(\cdot) = 0$  since  $\int_{\Omega} (\tilde{y}'(x) - y'(x)) \bar{y}(x) dx = 0$ . That is, among those functions satisfying (2.17), (2.16) admits a unique solution. Consequently,  $y'_\varepsilon(\cdot) \rightarrow y'(\cdot)$  weakly in  $W_0^{1,2}(\Omega)$  for the whole sequence  $\varepsilon \rightarrow 0^+$ .

Finally, by (2.18) and (2.16),

$$\begin{aligned} &\lim_{\varepsilon \rightarrow 0^+} \int_{\Omega} \langle \bar{a}(x) \nabla y'_\varepsilon(x), \nabla y'_\varepsilon(x) \rangle dx \\ &= \lim_{\varepsilon \rightarrow 0^+} \int_{\Omega} \left( \lambda_\varepsilon |y'_\varepsilon(x)|^2 + \lambda'_\varepsilon y'_\varepsilon(x) \bar{y}(x) + \langle ([a(x) - \bar{a}(x)] \nabla y_\varepsilon(x)), \nabla y'_\varepsilon(x) \rangle \right) dx \\ &= \int_{\Omega} (\bar{\lambda} |y'(x)|^2 + \lambda' y'(x) \bar{y}(x) + \langle ([a(x) - \bar{a}(x)] \nabla \bar{y}(x)), \nabla y'(x) \rangle) dx \\ &= \int_{\Omega} \langle \bar{a}(x) \nabla y'(x), \nabla y'(x) \rangle dx. \end{aligned}$$

Combining the above with the weak convergence of  $y'_\varepsilon(\cdot)$ , we get that  $y'_\varepsilon(\cdot) \rightarrow y'(\cdot)$  strongly in  $W_0^{1,2}(\Omega)$  for the whole sequence  $\varepsilon \rightarrow 0^+$ . This complete the proof.  $\square$

The following gives some general results of Problem  $(\bar{\Lambda}(\mathcal{A}))$  and Problem  $(\underline{\Lambda}(\mathcal{A}))$  associated with any convex and closed set  $\mathcal{A} \subseteq \mathcal{M}[\mu_0, \mu_1]$ , based on the above proposition.

**Corollary 2.6** *Let  $\Omega$  be a bounded Lipschitz domain. Let  $0 < \mu_0 \leq \mu_1 < \infty$ , and  $\mathcal{A} \subseteq \mathcal{M}[\mu_0, \mu_1]$  be convex and closed in  $L^1(\Omega; \mathbb{S}^n)$ . Then the following conclusions are true:*

(i) *Problem  $(\bar{\Lambda}(\mathcal{A}))$  admits an optimal solution. Further,  $\bar{a}(\cdot) \in \mathcal{A}$  is an optimal solution to Problem  $(\bar{\Lambda}(\mathcal{A}))$  if and only if*

$$\int_{\Omega} \langle [a(x) - \bar{a}(x)] \nabla y_{\bar{a}(\cdot)}(x), \nabla y_{\bar{a}(\cdot)}(x) \rangle dx \leq 0, \quad \forall a(\cdot) \in \mathcal{A}. \quad (2.21)$$

(ii) *If  $\underline{a}(\cdot) \in \mathcal{A}$  is an optimal solution to Problem  $(\underline{\Lambda}(\mathcal{A}))$ . Then*

$$\int_{\Omega} \langle [a(x) - \underline{a}(x)] \nabla y_{\underline{a}(\cdot)}(x), \nabla y_{\underline{a}(\cdot)}(x) \rangle dx \geq 0, \quad \forall a(\cdot) \in \mathcal{A}. \quad (2.22)$$

**Proof** (i) From Proposition 2.2 and Theorem 2.4, we know that  $a(\cdot) \mapsto \lambda_{a(\cdot)}$  is concave and continuous from  $\mathcal{A}$  to  $\mathbb{R}$ . Hence, a standard argument involving Mazur's Theorem applies to get the existence of an optimal solution to Problem  $(\bar{\Lambda}(\mathcal{A}))$ .

Next, if  $\bar{a}(\cdot) \in \mathcal{A}$  is a maximum of  $a(\cdot) \mapsto \lambda_{a(\cdot)}$ , then for any  $a(\cdot) \in \mathcal{A}$ , making use of the convexity of  $\mathcal{A}$ , we have

$$0 \geq \lim_{\varepsilon \downarrow 0} \frac{\lambda_{\bar{a}(\cdot)+\varepsilon[a(\cdot)-\bar{a}(\cdot)]} - \lambda_{\bar{a}(\cdot)}}{\varepsilon} = \lambda' = \int_{\Omega} \langle [a(x) - \bar{a}(x)] \nabla y_{\bar{a}(\cdot)}(x), \nabla y_{\bar{a}(\cdot)}(x) \rangle dx.$$

This gives (2.21). Conversely, suppose (2.21) holds. Then by the concavity of  $a(\cdot) \mapsto \lambda_{a(\cdot)}$ , we see that for any  $a(\cdot) \in \mathcal{A}$ ,  $\varepsilon \mapsto \lambda_{\bar{a}(\cdot)+\varepsilon[a(\cdot)-\bar{a}(\cdot)]}$  is concave as well. Thus,  $\varepsilon \mapsto \frac{d}{d\varepsilon} \lambda_{\bar{a}(\cdot)+\varepsilon[a(\cdot)-\bar{a}(\cdot)]}$  is non-increasing. Consequently,

$$\begin{aligned} \lambda_{a(\cdot)} - \lambda_{\bar{a}(\cdot)} &= \frac{1}{\varepsilon} \left[ (1 - \varepsilon) \lambda_{\bar{a}(\cdot)} + \varepsilon \lambda_{a(\cdot)} - \lambda_{\bar{a}(\cdot)} \right] \\ &\leq \frac{1}{\varepsilon} \left[ \lambda_{\bar{a}(\cdot)+\varepsilon[a(\cdot)-\bar{a}(\cdot)]} - \lambda_{\bar{a}(\cdot)} \right] \\ &\leq \left[ \frac{d}{d\varepsilon} \lambda_{\bar{a}(\cdot)+\varepsilon[a(\cdot)-\bar{a}(\cdot)]} \right]_{\varepsilon=0} \\ &= \int_{\Omega} \langle [a(x) - \bar{a}(x)] \nabla y_{\bar{a}(\cdot)}(x), \nabla y_{\bar{a}(\cdot)}(x) \rangle dx \leq 0. \end{aligned}$$

Hence,  $\bar{a}(\cdot)$  is a maximum of  $\lambda_{a(\cdot)}$  over  $\mathcal{A}$ .

(ii) From Proposition 2.5, we see that

$$\lambda_{\underline{a}(\cdot)+\varepsilon[a(\cdot)-\underline{a}(\cdot)]} = \lambda_{\underline{a}(\cdot)} + \varepsilon \int_{\Omega} \langle [a(x) - \underline{a}(x)] \nabla y_{\underline{a}(\cdot)}(x), \nabla y_{\underline{a}(\cdot)}(x) \rangle dx + o(\varepsilon).$$

Hence, if  $\underline{a}(\cdot) \in \mathcal{A}$  is a solution to Problem  $(\underline{\Lambda}(\mathcal{A}))$ , then (2.22) holds.  $\square$

We note that Corollary 2.6 part (i) gives the existence and characterization of an optimal solutions to Problem  $(\bar{\Lambda}(\mathcal{A}))$ , thanks to the concavity of the map  $a(\cdot) \mapsto \lambda_{a(\cdot)}$ . Whereas, part (ii) of Corollary 2.6 only gives a necessary condition for a possible solution of Problem  $(\underline{\Lambda}(\mathcal{A}))$ , and no existence of optimal solution is guaranteed.

### 3 A convexification of problem $(\bar{\Lambda}[\alpha, \beta])$

Let us return to Problem  $(\bar{\Lambda}[\alpha, \beta])$ . Since  $\mathcal{U}[\alpha, \beta]$  is not convex, the existence of an optimal solution is not guaranteed. In this section, we consider a convexification of Problem  $(\bar{\Lambda}[\alpha, \beta])$ .

For  $0 \leq \alpha \leq \beta \leq 1$ , we introduce the following:

$$\Sigma[\alpha, \beta] = \left\{ \sigma : \Omega \rightarrow [0, 1] \mid \sigma(\cdot) \text{ measurable, } \alpha|\Omega| \leq \int_{\Omega} \sigma(x) dx \leq \beta|\Omega| \right\}, \quad (3.1)$$

which is convex and closed in  $L^1(\Omega)$ . Recalling  $\mathcal{U}[\alpha, \beta]$  defined by (1.17), one has

$$\overline{\text{co}\{\mathcal{U}[\alpha, \beta]\}}^{L^1(\Omega; \mathbb{R})} = \Sigma[\alpha, \beta], \quad (3.2)$$

where the left hand side of the above is the closed convex hull of  $\mathcal{U}[\alpha, \beta]$  in  $L^1(\Omega; \mathbb{R})$ . Now, for given  $0 \leq \alpha \leq \beta \leq 1$ , and  $A_0, A_1 \in M[\mu_0, \mu_1]$  with  $0 < \mu_0 \leq \mu_1 < \infty$  such that (1.24) holds, with  $A(\cdot)$  defined by (1.16), one sees that

$$A(\Sigma[\alpha, \beta]) = \left\{ A_0 + \sigma(\cdot)(A_1 - A_0) \mid \sigma(\cdot) \in \Sigma[\alpha, \beta] \right\}$$

is convex and closed in  $L^1(\Omega; \mathbb{R})$ . For any  $\sigma(\cdot) \in \Sigma[\alpha, \beta]$ , we consider the following state equation

$$\begin{cases} -\nabla \cdot (A(\sigma(x)) \nabla y(x)) = \lambda y(x), & x \in \Omega, \\ y|_{\partial\Omega} = 0, \end{cases} \quad (3.3)$$

where  $A(\sigma) \equiv A_0 + \sigma(A_1 - A_0)$ . Denote the corresponding principal eigenvalue and normalized principal eigenfunction by

$$\lambda_{\sigma(\cdot)} \equiv \lambda_{A_0 + \sigma(\cdot)(A_1 - A_0)}, \quad y_{\sigma(\cdot)} = y_{A_0 + \sigma(\cdot)(A_1 - A_0)}(\cdot),$$

respectively. We pose the following convexified problem.

**Problem** ( $\bar{\Lambda}^c[\alpha, \beta]$ ). Let  $0 \leq \alpha \leq \beta \leq 1$ ,  $0 < \mu_0 \leq \mu_1 < \infty$ , and  $A_0, A_1 \in M[\mu_0, \mu_1]$  satisfying (1.24). Find a  $\bar{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$  such that

$$\lambda_{\bar{\sigma}(\cdot)} = \sup_{\sigma(\cdot) \in \Sigma[\alpha, \beta]} \lambda_{\sigma(\cdot)}. \quad (3.4)$$

Any  $\bar{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$  satisfying (3.4) is called an *optimal control* of Problem ( $\bar{\Lambda}^c[\alpha, \beta]$ ), which is also called an *optimal convexified relaxed control* of Problem ( $\bar{\Lambda}[\alpha, \beta]$ ). The superscript “ $c$ ” in  $\bar{\Lambda}^c[\alpha, \beta]$  indicates the “convexification”. Note that if  $\bar{\sigma}(\cdot)$  is an optimal control of Problem ( $\bar{\Lambda}^c[\alpha, \beta]$ ) and

$$|(0 < \bar{\sigma}(\cdot) < 1)| = 0, \quad (3.5)$$

where

$$(0 < \bar{\sigma}(\cdot) < 1) = \{x \in \Omega \mid 0 < \bar{\sigma}(x) < 1\},$$

then

$$\bar{u}(\cdot) = \chi_{\Omega_1}(\cdot), \quad \Omega_1 = \{\bar{\sigma}(\cdot) = 1\}$$

is an optimal control of Problem ( $\bar{\Lambda}[\alpha, \beta]$ ). The following result gives the necessary conditions when (3.5) fails.

**Theorem 3.1** Problem ( $\bar{\Lambda}^c[\alpha, \beta]$ ) admits an optimal control  $\bar{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$ . Suppose

$$|(0 < \bar{\sigma}(\cdot) < 1)| > 0, \quad (3.6)$$

and  $\bar{y}(\cdot)$  is the corresponding optimal state. Then

$$\begin{cases} \langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \equiv C, & \text{a.e. } x \in (0 < \bar{\sigma}(\cdot) < 1), \\ \langle (A_1 - A_0) \nabla \bar{y}(x'), \nabla \bar{y}(x') \rangle \leq \langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \leq \langle (A_1 - A_0) \nabla \bar{y}(x''), \nabla \bar{y}(x'') \rangle, \\ \quad \text{a.e. } x' \in (\bar{\sigma}(\cdot) = 0), \quad \text{a.e. } x \in (0 < \bar{\sigma}(\cdot) < 1), \quad \text{a.e. } x'' \in (\bar{\sigma}(\cdot) = 1). \end{cases} \quad (3.7)$$

Further, in the case

$$\int_{\Omega} \bar{\sigma}(x) dx < \beta |\Omega|, \quad (3.8)$$

the following holds:

$$\langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \leq 0, \quad \text{a.e. } x \in (0 \leq \bar{\sigma}(\cdot) < 1); \quad (3.9)$$

in the case

$$\int_{\Omega} \bar{\sigma}(x) dx > \alpha |\Omega|, \quad (3.10)$$

the following holds:

$$\langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \geq 0, \quad \text{a.e.} x \in (0 < \bar{\sigma}(\cdot) \leq 1); \quad (3.11)$$

and in the case

$$\alpha |\Omega| < \int_{\Omega} \bar{\sigma}(x) dx < \beta |\Omega|, \quad (3.12)$$

the following holds:

$$\langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle = 0, \quad \text{a.e.} x \in (0 < \bar{\sigma}(\cdot) < 1). \quad (3.13)$$

**Proof** Since  $A(\Sigma[\alpha, \beta])$  is convex and closed in  $L^1(\Omega; \mathbb{S}^n)$ , by Corollary 2.6 part (i), we have that Problem  $(\bar{\Lambda}^c[\alpha, \beta])$  admits an optimal solution  $\bar{\sigma}(\cdot)$  and the following is its characterization:

$$0 \geq \int_{\Omega} (\sigma(x) - \bar{\sigma}(x)) \langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle dx, \quad \forall \sigma(\cdot) \in \Sigma[\alpha, \beta]. \quad (3.14)$$

We now look at further necessary conditions for  $(\bar{y}(\cdot), \bar{\sigma}(\cdot))$ .

Since (3.6) holds, for any  $v(\cdot) \in \mathcal{V}_0$  where

$$\mathcal{V}_0 \equiv \left\{ v(\cdot) \in L^{\infty}(\Omega; \mathbb{R}) \mid \int_{\Omega} v(x) dx = 0 \right\},$$

with

$$\text{supp } v(\cdot) \subseteq (0 < \bar{\sigma}(\cdot) < 1),$$

one has  $\sigma(\cdot) = \bar{\sigma}(\cdot) \pm \varepsilon v(\cdot) \in \Sigma[\alpha, \beta]$  as long as  $\varepsilon > 0$  is small enough. By taking such a  $\sigma(\cdot)$  in (3.14), we have

$$\int_{\Omega} v(x) \langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle dx = 0, \quad \forall v(\cdot) \in \mathcal{V}_0. \quad (3.15)$$

This leads to the first identity in (3.7).

Next, let the set  $(\bar{\sigma}(\cdot) = 0)$  have a positive measure. Then take any  $v(\cdot) \in L^{\infty}(\Omega)$  with  $v(x) \geq 0$ , supported on  $(\bar{\sigma}(\cdot) = 0)$ , and any  $w(\cdot) \in L^{\infty}(\Omega)$  with  $w(x) \geq 0$ , supported on  $(0 < \bar{\sigma}(\cdot) \leq 1)$  (which has a positive measure by (3.6)), and

$$\int_{(0 < \bar{\sigma}(\cdot) \leq 1)} w(x) dx = \int_{(\bar{\sigma}(\cdot) = 0)} v(x) dx. \quad (3.16)$$

Then for  $\varepsilon > 0$  small enough,

$$\sigma(\cdot) = \bar{\sigma}(\cdot) + \varepsilon v(\cdot) - \varepsilon w(\cdot) \in \Sigma[\alpha, \beta].$$

Hence, using such a  $\sigma(\cdot)$  in (3.14), one obtains the first inequality in the second conclusion of (3.7). Likewise, we can obtain the second inequality in the second conclusion of (3.7).

Further, if (3.8) holds, we may take  $v(\cdot) \in L^{\infty}(\Omega)$  with  $v(\cdot) \geq 0$ , supported in  $(0 \leq \bar{\sigma}(\cdot) < 1)$ . Then, for  $\varepsilon > 0$  small,  $\bar{\sigma}(\cdot) + \varepsilon v(\cdot) \in \Sigma[\alpha, \beta]$ . Taking such a  $\sigma(\cdot)$  in (3.14), we obtain (3.9). Likewise we can obtain (3.11) under (3.10). Finally, combining the above two cases, we obtain (3.13) under (3.12). This completes the proof.  $\square$

We now present an interesting corollary.

**Corollary 3.2** Let  $0 < \alpha < \beta < 1$ , and  $A_0 < A_1$ . Then there exists an optimal control  $\bar{\sigma}(\cdot)$  of Problem  $(\bar{\Lambda}^c[\alpha, \beta])$  such that

$$\int_{\Omega} \bar{\sigma}(x) dx = \beta |\Omega|. \quad (3.17)$$

Further, if  $\bar{\sigma}(\cdot)$  is a constant on  $\Omega$ , then  $\bar{\sigma}(\cdot)$  is not an optimal control of Problem  $(\bar{\Lambda}^c[\alpha, \beta])$ .

**Proof** The first conclusion is obvious. It follows easily from the monotonicity of the principal eigen map (see (1.10)).

Next, suppose  $\bar{\sigma}(x) \equiv \sigma_0 \in [\alpha, \beta] \subset (0, 1)$  (so that  $\bar{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$ ). If such a  $\bar{\sigma}(\cdot)$  is optimal, then by the first equation in (3.7), we have

$$\langle (A_1 - A_0) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \equiv C, \quad \text{a.e. } x \in \Omega. \quad (3.18)$$

On the other hand, since  $A(\bar{\sigma}(x))$  is a constant matrix,  $\bar{y}(\cdot) \in C^{\infty}(\Omega)$ . Thus making use of the fact that  $\bar{y}(\cdot)$  is strictly positive in  $\Omega$  and is zero on  $\partial\Omega$ , one sees that  $\bar{y}(\cdot)$  attains its maximum on  $\bar{\Omega}$  at some point  $x_0 \in \Omega$ . This implies  $\nabla \bar{y}(x_0) = 0$  and therefore,

$$C = \langle (A_1 - A_0) \nabla \bar{y}(x_0), \nabla \bar{y}(x_0) \rangle = 0.$$

Since  $A_1 - A_0$  is positive definite,  $C = 0$  and (3.18) implies  $\bar{y}(\cdot) \equiv 0$ . This is a contradiction since  $\bar{y}(\cdot)$  is positive in  $\Omega$ . Hence, such a  $\bar{\sigma}(\cdot)$  is not optimal.  $\square$

Note that a constant  $\bar{\sigma}(\cdot) \equiv \sigma_0$  is not an optimal control means that a perfect mixture of two different materials does not give the optimal solution to the problem.

To conclude this section, we present an illustrative example for the case  $\alpha = 0, \beta = 1$  with both  $A_0 \leq A_1$  and  $A_0 \geq A_1$  fail.

**Example 3.3** Let  $\Omega = [-1, 1] \times [-1, 1]$  which is a bounded Lipschitz domain,  $\alpha = 0, \beta = 1$ , and

$$A_0 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{3}{2} \end{pmatrix}, \quad A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus, both  $A_0 \leq A_1$  and  $A_0 \geq A_1$  fail. Let us consider the following problem:

$$\begin{cases} -(ay_{x_1 x_1} + by_{x_2 x_2}) = \lambda y, & \text{in } \Omega, \\ y|_{\partial\Omega} = 0, \end{cases}$$

for any given  $a, b > 0$ . Then we can check directly that the principal eigen pair is given by

$$\begin{cases} \lambda = \frac{(a+b)\pi^2}{4}, \\ y(x_1, x_2) = \cos \frac{\pi x_1}{2} \cos \frac{\pi x_2}{2}, \quad (x_1, x_2) \in [-1, 1] \times [-1, 1]. \end{cases}$$

From this, we see that for any constant  $\sigma \in [0, 1]$ , one has

$$A_0 + \sigma(A_1 - A_0) = \begin{pmatrix} \frac{1}{2} + \frac{\sigma}{2} & 0 \\ 0 & \frac{3}{2} - \frac{\sigma}{2} \end{pmatrix},$$

and by the above calculation,

$$\lambda_{A_0 + \sigma(A_1 - A_0)} = \frac{\pi^2}{2} = \lambda_{A_0} = \lambda_{A_1}, \quad \forall \sigma \in [0, 1]. \quad (3.19)$$

This shows that the map  $a(\cdot) \mapsto \lambda_{a(\cdot)}$  is not strictly convex. If  $\sigma(x) \equiv \sigma \in (0, 1)$  is optimal, then by Theorem 3.1, we should have

$$C \equiv \langle (A_1 - A_0) \nabla y(x), \nabla y(x) \rangle = \frac{\pi^2}{4} \langle \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} \sin \frac{\pi x_1}{2} \cos \frac{\pi x_2}{2} \\ \cos \frac{\pi x_1}{2} \sin \frac{\pi x_2}{2} \end{pmatrix}, \begin{pmatrix} \sin \frac{\pi x_1}{2} \cos \frac{\pi x_2}{2} \\ \cos \frac{\pi x_1}{2} \sin \frac{\pi x_2}{2} \end{pmatrix} \rangle \\ = \frac{\pi^2}{8} \left[ \sin^2 \frac{\pi x_1}{2} \cos^2 \frac{\pi x_2}{2} - \cos^2 \frac{\pi x_1}{2} \sin^2 \frac{\pi x_2}{2} \right] = \frac{\pi^2}{8} \sin \frac{\pi(x_1 + x_2)}{2} \sin \frac{\pi(x_1 - x_2)}{2},$$

which is impossible. Hence,  $\sigma(x) \equiv \sigma \in (0, 1)$  is not an optimal solution to the corresponding maximization problem. Due to (3.19), we see that both  $\sigma(\cdot) = 0$  and  $\sigma(\cdot) = 1$  are not optimal either. By the way, the above also roughly means that in the current case, if  $A_0$  and  $A_1$  represent the heat diffusibility of the two materials, then the uniform mixture of any ratio of these two materials is not optimal for Problem  $(\bar{\Lambda}^c[0, 1])$ . It is not clear to us at the moment what is an optimal control for this problem.

## 4 Relaxation of problem $(\underline{\Lambda}[\alpha, \beta])$

Mimicking Problem  $(\bar{\Lambda}^c[\alpha, \beta])$ , we may pose the following problem.

**Problem  $(\underline{\Lambda}^c[\alpha, \beta])$ .** Let  $0 < \mu_0 \leq \mu_1 < \infty$ ,  $A_0, A_1 \in M[\mu_0, \mu_1]$  and  $0 \leq \alpha \leq \beta \leq 1$  satisfying (1.24). Find a  $\underline{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$  such that

$$\lambda_{\underline{\sigma}(\cdot)} = \inf_{\sigma(\cdot) \in \Sigma[\alpha, \beta]} \lambda_{\sigma(\cdot)}. \quad (4.1)$$

Note that although  $\Sigma[\alpha, \beta]$  is convex and closed (in  $L^1(\Omega; \mathbb{R})$ ),  $A(\cdot) \mapsto \lambda_{A(\cdot)}$  is concave and not necessarily convex in general. Therefore, it is not clear if the map  $\sigma(\cdot) \mapsto \lambda_{\sigma(\cdot)}$  admits a minimum on  $\Sigma[\alpha, \beta]$ . In another word, the above Problem  $(\underline{\Lambda}^c[\alpha, \beta])$  might not admit a minimum over  $\Sigma[\alpha, \beta]$  in general. Hence, instead of Problem  $(\underline{\Lambda}^c[\alpha, \beta])$ , we will introduce another relaxation of Problem  $(\underline{\Lambda}[\alpha, \beta])$ , for which the existence of an optimal solution is guaranteed. To this end, let us recall some results relevant to the so-called  $H$ -convergence, which will play an essential role in the relaxation of Problem  $(\underline{\Lambda}[\alpha, \beta])$ .

### 4.1 $H$ -convergence

We recall the following definition.

**Definition 4.1** A sequence  $\{a_\varepsilon(\cdot)\} \subseteq \mathcal{M}[\mu_0, \mu_1]$  is said to be  $H$ -convergent to  $a^*(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$  on  $\Omega$ , denoted by  $a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot)$ , if for any  $f \in W^{-1,2}(\Omega)$ , the weak solution  $y_\varepsilon(\cdot)$  of the following problem

$$\begin{cases} -\nabla \cdot (a_\varepsilon(x) \nabla y_\varepsilon(x)) = f, & \text{in } \Omega, \\ y_\varepsilon|_{\partial\Omega} = 0 \end{cases} \quad (4.2)$$

has the property that

$$y_\varepsilon(\cdot) \rightarrow y^*(\cdot), \quad \text{weakly in } W_0^{1,2}(\Omega) \quad (4.3)$$

with  $y^*(\cdot)$  being the weak solution to the following:

$$\begin{cases} -\nabla \cdot (a^*(x) \nabla y^*(x)) = f, & \text{in } \Omega, \\ y^*|_{\partial\Omega} = 0. \end{cases} \quad (4.4)$$

In 1968, Spagnolo ([31]) introduced the above notion, called the *G-convergence*, for symmetric operators (i.e., each  $a_\varepsilon(\cdot)$  is symmetric matrix valued and so is  $a^*(\cdot)$ ). The notion was generalized by Tartar for possibly non-symmetric operators ([32]), and is called the *H-convergence*, for which the following additional condition is required:

$$a_\varepsilon(\cdot) \nabla y_\varepsilon(\cdot) \rightarrow a^*(\cdot) \nabla y^*(\cdot), \quad \text{weakly in } L^2(\Omega; \mathbb{R}^n), \quad (4.5)$$

which is automatically true when  $a_\varepsilon(\cdot)$  is symmetric and is *H*-convergent. It is known that for symmetric operators, the *G*-convergence is equivalent to the *H*-convergence ([3]). In the problems that we are studying, all the involved second order differential operators are symmetric. Hence, *G*-convergence will be enough. However, we prefer to use the name *H*-convergence instead, just keep in mind that we are treating the case of symmetric operators. For more words on *H*-convergence, see also [4, 17, 30].

Note that in the definition, the *H*-limit  $a^*(\cdot)$  of  $a_\varepsilon(\cdot)$  is independent of the choice of  $f \in W^{-1,2}(\Omega)$ , and the whole sequence (not just a subsequence)  $y_\varepsilon(\cdot)$  is required to be weakly convergent in  $W_0^{1,2}(\Omega)$ .

For any  $X \subseteq \mathcal{M}[\mu_0, \mu_1]$ , we denote the *H*-closure of  $X$  by  $\overline{X}^H$ . Let us now list some useful properties of *H*-convergence and *H*-closure, found in [3].

- (i) *Sequential compactness* The set  $\mathcal{M}[\mu_0, \mu_1]$  is sequentially compact under *H*-convergence, i.e., for any sequence  $\{a_k(\cdot)\}_{k \geq 1} \subseteq \mathcal{M}[\mu_0, \mu_1]$ , there exists a subsequence  $\{a_{k_i}(\cdot)\}_{i \geq 1}$  and an  $a^*(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$  such that  $a_{k_i}(\cdot) \xrightarrow{H} a^*(\cdot)$ .
- (ii) *Locality* If  $a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot)$  on  $\Omega$ , then  $a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot)$  on any subdomain  $\omega \subseteq \Omega$ .
- (iii) *Monotonicity and uniqueness* If  $a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot)$ ,  $b_\varepsilon(\cdot) \xrightarrow{H} b^*(\cdot)$  with  $a_\varepsilon(\cdot) \leq b_\varepsilon(\cdot)$ , then  $a^*(\cdot) \leq b^*(\cdot)$ . In particular, if  $0 < \mu_0 \leq a_\varepsilon(x) \leq \mu_1$ , then

$$\mu_0 \leq a^*(x) \leq \mu_1, \quad x \in \Omega.$$

Also, by taking  $b_\varepsilon(\cdot) = a_\varepsilon(\cdot)$ , one has

$$\left. \begin{array}{l} a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot) \\ a_\varepsilon(\cdot) \xrightarrow{H} b^*(\cdot) \end{array} \right\} \Rightarrow a^*(\cdot) = b^*(\cdot).$$

That is, the *H*-limit of a sequence is unique.

- (iv) *Non-homogeneous boundary conditions* Let

$$a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot), \quad \varphi \in W^{1,2}(\Omega), \quad f \in W^{-1,2}(\Omega),$$

and  $y_\varepsilon(\cdot)$  and  $y^*(\cdot)$  respectively be the solutions to the following:

$$\begin{cases} -\nabla \cdot (a_\varepsilon(x) \nabla y_\varepsilon(x)) = f, & \text{in } \Omega, \\ y_\varepsilon|_{\partial\Omega} = \varphi, \end{cases} \quad (4.6)$$

$$\begin{cases} -\nabla \cdot (a^*(x) \nabla y^*(x)) = f, & \text{in } \Omega, \\ y^*|_{\partial\Omega} = \varphi. \end{cases} \quad (4.7)$$

Then

$$y_\varepsilon(\cdot) \rightarrow y^*(\cdot), \quad \text{weakly in } W^{1,2}(\Omega).$$

- (v) *Metrizability* Let  $\{\mathbf{f}_\ell(\cdot)\}_{\ell \geq 1} \subseteq L^2(\Omega; \mathbb{R}^n)$  such that  $\{\nabla \cdot \mathbf{f}_\ell\}_{\ell \geq 1}$  is dense in  $W^{-1,2}(\Omega)$ . For any  $a(\cdot), b(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ , let  $y_\ell^{a(\cdot)}(\cdot)$  and  $y_\ell^{b(\cdot)}(\cdot)$  be the unique weak solutions

to the following:

$$\begin{cases} -\nabla \cdot (a(x) \nabla y_\ell^{a(\cdot)}(x)) = \nabla \cdot \mathbf{f}_\ell, & \text{in } \Omega, \\ y_\ell^{a(\cdot)}|_{\partial\Omega} = 0, \end{cases}$$

and

$$\begin{cases} -\nabla \cdot (b(x) \nabla y_\ell^{b(\cdot)}(x)) = \nabla \cdot \mathbf{f}_\ell, & \text{in } \Omega, \\ y_\ell^{b(\cdot)}|_{\partial\Omega} = 0. \end{cases}$$

Define

$$\rho(a(\cdot), b(\cdot)) = \sum_{\ell \geq 1} 2^{-\ell} \frac{\|y_\ell^{a(\cdot)}(\cdot) - y_\ell^{b(\cdot)}(\cdot)\|_2 + \|a(\cdot) \nabla y_\ell^{a(\cdot)}(\cdot) - b(\cdot) \nabla y_\ell^{b(\cdot)}(\cdot)\|_2}{\|\mathbf{f}_\ell\|_2}. \quad (4.8)$$

Then  $\rho(\cdot, \cdot)$  is a metric on  $\mathcal{M}[\mu_0, \mu_1]$  such that for any  $a_\varepsilon(\cdot), a^*(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ ,

$$a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot) \iff \rho(a_\varepsilon(\cdot), a^*(\cdot)) \rightarrow 0. \quad (4.9)$$

(vi) *Upper and lower bounds* Let

$$\begin{cases} a_\varepsilon(\cdot) \rightarrow \bar{a}(\cdot), & a_\varepsilon(\cdot)^{-1} \rightarrow \underline{a}(\cdot)^{-1}, \quad \text{weak* in } L^\infty(\Omega; \mathbb{S}^n), \\ a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot). \end{cases}$$

Then

$$\underline{a}(x) \leq a^*(x) \leq \bar{a}(x), \quad \text{a.e. } x \in \Omega. \quad (4.10)$$

(vii) *Commutativity with congruent transformation* Let  $Q \in \mathbb{R}^{n \times n}$  be non-singular. Then

$$a_\varepsilon(\cdot) \xrightarrow{H} a^*(\cdot) \iff Q a_\varepsilon(\cdot) Q^T \xrightarrow{H} Q a^*(\cdot) Q^T.$$

(viii) *Pointwiseness* Let  $G \subseteq M[\mu_0, \mu_1]$ , denote

$$L^\infty(\Omega; G) = \left\{ a(\cdot) \in L^\infty(\Omega; \mathbb{S}^n) \mid a(x) \in G, \text{ a.e. } x \in \Omega \right\},$$

and define

$$\overline{G}^H = \left\{ A \in M[\mu_0, \mu_1] \mid \chi_\Omega(\cdot) A \in \overline{L^\infty(\Omega; G)}^H \right\}.$$

Then

$$\overline{L^\infty(\Omega; G)}^H = L^\infty(\Omega; \overline{G}^H) \equiv \left\{ a^*(\cdot) \in \mathcal{M}[\mu_0, \mu_1] \mid a^*(x) \in \overline{G}^H, \text{ a.e. } x \in \Omega \right\}. \quad (4.11)$$

Namely,  $a^*(\cdot) \in \overline{L^\infty(\Omega; G)}^H$  if and only if for almost all  $x \in \Omega$ , there exists a sequence  $\{a_k(\cdot; x)\}_{k \geq 1} \subseteq L^\infty(\Omega; G)$  (depending on  $x$ ) such that

$$a_k(\cdot; x) \xrightarrow{H} \chi_\Omega(\cdot) a^*(x), \quad k \rightarrow \infty.$$

More generally, let  $\mathcal{Q} \subseteq \mathcal{M}[\mu_0, \mu_1]$  and define

$$\mathcal{Q}_x = \{q(x) \mid q(\cdot) \in \mathcal{Q}\}, \quad \forall x \in \Omega.$$

Then, under some mild conditions (see Theorem 2.3 in [20])

$$\overline{\mathcal{Q}}^H = \left\{ a^*(\cdot) \in \mathcal{M}[\mu_0, \mu_1] \mid a^*(x) \in \overline{\mathcal{Q}}_x^H, \text{ a.e. } x \in \Omega \right\}.$$

Note that by taking  $a_k(\cdot) = \chi_{\Omega}(\cdot)A$  with  $A \in G$ , we see that

$$G \subseteq \overline{G}^H. \quad (4.12)$$

We will see that  $G$  is a proper subset of  $\overline{G}^H$  below.

## 4.2 Lamination

In this subsection, we consider a special case involving two matrices, which will be useful in our relaxation of Problem  $(\Lambda[\alpha, \beta])$ . Let us first present the following result.

**Theorem 4.2** *Let  $0 < \mu_0 \leq \mu_1 < \infty$  and  $A, B \in M[\mu_0, \mu_1]$  be fixed.*

(i) *For any  $\theta \in (0, 1)$  and  $e \in S^{n-1} \equiv \{x \in \mathbb{R}^n \mid |x| = 1\}$ , define*

$$\mathcal{H}_\varepsilon(A, B; \theta, e) = \begin{cases} A, & \left\{ \frac{\langle x, e \rangle}{\varepsilon} \right\} \in [\theta, 1), \\ B, & \left\{ \frac{\langle x, e \rangle}{\varepsilon} \right\} \in [0, \theta), \end{cases} \quad x \in \mathbb{R}^n, \quad (4.13)$$

where  $\{r\} = r - [r]$  is the decimal part of the real number  $r$ . Then, as  $\varepsilon \rightarrow 0^+$ ,

$$\begin{aligned} \mathcal{H}_\varepsilon(A, B; \theta, e) &\xrightarrow{H} \mathcal{H}[A, B; \theta, e] \\ &\equiv (1 - \theta)A + \theta B - \frac{\theta(1 - \theta)(A - B)ee^T(A - B)}{e^T[\theta A + (1 - \theta)B]e} \\ &= A - \theta(A - B) - \frac{\theta(1 - \theta)(A - B)ee^T(A - B)}{e^T[B + \theta(A - B)]e} \in \overline{\{A, B\}}^H \subseteq M[\mu_0, \mu_1]. \end{aligned} \quad (4.14)$$

(ii) *For any  $m \geq 1$ , let*

$$\begin{aligned} \Gamma^m(A, B) &= \left\{ A^* \in \mathbb{S}^n \mid (1 - \theta)(A - B) = (A^* - B) \left[ I + \theta \sum_{k=1}^m \beta_k \frac{e_k e_k^T (A - B)}{e_k^T B e_k} \right], \right. \\ &\quad \left. \text{for some } \theta \in [0, 1], e_k \in S^{n-1}, \beta_k \geq 0, \sum_{k=1}^m \beta_k = 1 \right\}, \end{aligned} \quad (4.15)$$

and

$$\Gamma(A, B) = \bigcup_{m=1}^{\infty} \Gamma^m(A, B) \subseteq \overline{\{A, B\}}^H. \quad (4.16)$$

Then

$$\Gamma^1(A, B) = \left\{ \mathcal{H}[A, B; \theta, e] \mid \theta \in [0, 1], e \in S^{n-1} \right\} \equiv \mathcal{H}[A, B; [0, 1], S^{n-1}], \quad (4.17)$$

and

$$\Gamma^1(\Gamma(A, B), B) \subseteq \Gamma(A, B). \quad (4.18)$$

(iii) *For any  $\theta \in [0, 1]$ ,  $H \in \mathbb{S}^n$  with  $H \geq 0$  and  $\text{tr}(H) = 1$ , the matrix*

$$I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B)$$

is non-singular, and

$$\Gamma(A, B) = \left\{ B + (1 - \theta)(A - B) \left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right]^{-1} \mid \theta \in [0, 1], H \geq 0, \text{tr}(H) = 1 \right\}. \quad (4.19)$$

### Proof

- (i) Relation (4.14) follows from [3], Corollary 1.3.34 (see [24] also).
- (ii) This is a restatement of Lemma 2.2.3 of [3].
- (iii) Let  $\theta \in (0, 1)$  and  $H > 0$ ,  $\text{tr}(H) = 1$ , we have  $I - H \geq 0$ , and

$$\begin{aligned} & \det \left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right] \\ &= \det(B^{\frac{1}{2}}) \det \left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} A - \theta B^{-\frac{1}{2}} H B^{\frac{1}{2}} \right] \det(B^{-\frac{1}{2}}) \\ &= \det \left[ I + \theta H (B^{-\frac{1}{2}} A B^{-\frac{1}{2}} - I) \right] \\ &= \det(H^{-\frac{1}{2}}) \det \left[ I - H + (1 - \theta)H + \theta H B^{-\frac{1}{2}} A B^{-\frac{1}{2}} \right] \det(H^{\frac{1}{2}}) \\ &= \det \left[ I - H + (1 - \theta)H + \theta H^{\frac{1}{2}} B^{-\frac{1}{2}} A B^{-\frac{1}{2}} H^{\frac{1}{2}} \right] \\ &\geq \det \left[ I - H + (1 - \theta)H + \theta \frac{\mu_0}{\mu_1} H \right] = \det \left[ I - \theta \frac{\mu_1 - \mu_0}{\mu_1} H \right] \\ &= \det \left[ \frac{\theta \mu_0 + (1 - \theta)\mu_1}{\mu_1} I + \frac{\theta(\mu_1 - \mu_0)}{\mu_1} (I - H) \right] \geq \left[ \frac{\theta \mu_0 + (1 - \theta)\mu_1}{\mu_1} \right]^n. \end{aligned}$$

By continuity, for any  $\theta \in [0, 1]$ ,  $H \geq 0$  (instead of just  $H > 0$ ), with  $\text{tr } H = 1$ , we have

$$\det \left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right] \geq \left[ \frac{\theta \mu_0 + (1 - \theta)\mu_1}{\mu_1} \right]^n > 0.$$

Therefore,  $\left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right]$  is non-singular.

On the other hand, if  $A^* \in \Gamma(A, B)$ , then

$$(1 - \theta)(A - B) = (A^* - B) \left[ I + \theta \sum_{k=1}^m \beta_k \frac{e_k e_k^T (A - B)}{e_k^T B e_k} \right], \quad (4.20)$$

for some  $\theta \in [0, 1]$ ,  $\beta_k \geq 0$ ,  $\sum_{k=1}^m \beta_k = 1$ , and  $e_k \in S^{n-1}$ . Note that

$$\sum_{k=1}^m \beta_k \frac{e_k e_k^T}{e_k^T B e_k} = B^{-\frac{1}{2}} \sum_{k=1}^m \beta_k \frac{(B^{\frac{1}{2}} e_k)}{|B^{\frac{1}{2}} e_k|} \left( \frac{(B^{\frac{1}{2}} e_k)}{|B^{\frac{1}{2}} e_k|} \right)^T B^{-\frac{1}{2}} \equiv B^{-\frac{1}{2}} H B^{-\frac{1}{2}},$$

where

$$H = \sum_{k=1}^m \beta_k \frac{(B^{\frac{1}{2}} e_k)^T}{|B^{\frac{1}{2}} e_k|} \left( \frac{(B^{\frac{1}{2}} e_k)}{|B^{\frac{1}{2}} e_k|} \right)^T \geq 0, \quad \text{tr}(H) = \sum_{k=1}^m \beta_k = 1.$$

Thus, (4.20) is equivalent to the following:

$$(1 - \theta)(A - B) = (A^* - B) \left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right]. \quad (4.21)$$

Then by the invertibility of  $\left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right]$ , we have

$$A^* = B + (1 - \theta)(A - B) \left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right]^{-1}. \quad (4.22)$$

Conversely, if (4.22) holds for some  $\theta \in [0, 1]$ , and  $H \geq 0$ ,  $\text{tr}(H) = 1$ , then (4.21) holds. Moreover, it is easy to see that

$$H = \sum_{k=1}^n \beta_k \xi_k \xi_k^T$$

with  $\beta_k \geq 0$  ( $1 \leq k \leq n$ ),  $\sum_{k=1}^n \beta_k = 1$  and  $\xi_k \in S^{n-1}$  ( $1 \leq k \leq n$ ). Thus

$$B^{-\frac{1}{2}} H B^{-\frac{1}{2}} = \sum_{k=1}^n \beta_k \frac{\xi_k \xi_k^T}{e_k^T B e_k},$$

where

$$e_k = \frac{B^{-\frac{1}{2}} \xi_k}{|B^{-\frac{1}{2}} \xi_k|} \in S^{n-1}, \quad 1 \leq k \leq n.$$

Hence,

$$\begin{aligned} (1 - \theta)(A - B) &= (A^* - B) \left[ I + \theta B^{-\frac{1}{2}} \sum_{k=1}^n \beta_k \frac{\xi_k \xi_k^T}{e_k^T B e_k} B^{-\frac{1}{2}} (A - B) \right] \\ &= (A^* - B) \left[ I + \theta \sum_{k=1}^n \beta_k \frac{e_k e_k^T}{e_k^T B e_k} (A - B) \right], \end{aligned}$$

which means  $A^* \in \Gamma(A, B)$ . This completes the proof.  $\square$

In the above, any element in  $\Gamma(A, B)$  is called a *lamination* of  $A$  with base  $B$ . From (4.19), we see that  $A, B \in \Gamma(A, B)$  (by taking  $\theta = 0, 1$ ). Thus,

$$\{A, B\} \subsetneq \Gamma(A, B) \subseteq \overline{\{A, B\}}^H.$$

Note also that for any  $\theta \in (0, 1)$ ,  $H \geq 0$ ,  $\text{tr}(H) = 1$  it holds that (see (4.10))

$$\begin{aligned} \left( (1 - \theta)A^{-1} + \theta B^{-1} \right)^{-1} &\leq B + (1 - \theta)(A - B) \left[ I + \theta B^{-\frac{1}{2}} H B^{-\frac{1}{2}} (A - B) \right]^{-1} \\ &\leq (1 - \theta)A + \theta B. \end{aligned} \quad (4.23)$$

This gives bounds for elements in  $\Gamma(A, B) \subseteq M[\mu_0, \mu_1]$ . Further, we should keep in mind some facts about the set  $\Gamma(A, B)$ :

- It is possible that  $\Gamma(A, B) \neq \Gamma(B, A)$  ( $n \geq 3$ )
- $\Gamma(A, B)$  is not necessarily convex, and even  $\overline{\{A, B\}}^H$  might be non-convex.
- It is possible that  $\overline{\{A, B\}}^H \neq \Gamma(A, B) \cup \Gamma(B, A)$ .
- Even for  $A = \lambda B$  with  $B$  being diagonal ( $n \geq 2$  and  $B \neq \gamma I$  for any  $\gamma \in \mathbb{R}$ , of course), as long as  $\lambda \neq 1$ ,  $\Gamma(A, B)$  contains non-diagonal matrices.

### 4.3 Relaxation problem

In this subsection, we fix  $0 < \mu_0 \leq \mu_1 < \infty$ ,  $0 \leq \alpha \leq \beta \leq 1$ , and  $G = \{A_0, A_1\} \subseteq M[\mu_0, \mu_1]$  satisfying (1.24). For a domain  $\omega$ , denote

$$\mathcal{A}_\omega[\alpha, \beta] = \left\{ A_0 + \chi_{\Omega_1}(\cdot)(A_1 - A_0) \mid \Omega_1 \subseteq \omega \text{ measurable, } \alpha|\omega| \leq |\Omega_1| \leq \beta|\omega| \right\}.$$

Recall (see (1.18)) that  $\mathcal{A}[\alpha, \beta] = \mathcal{A}_\Omega[\alpha, \beta]$ . We first present a simple result.

**Proposition 4.3** *Assume  $A_0 \neq A_1$ . The following hold:*

$$\alpha = 0 \iff \chi_\Omega(\cdot)A_0 \in \overline{\mathcal{A}[\alpha, \beta]}^H, \quad (4.24)$$

$$\beta = 1 \iff \chi_\Omega(\cdot)A_1 \in \overline{\mathcal{A}[\alpha, \beta]}^H. \quad (4.25)$$

**Proof** We just prove (4.24). One can prove (4.25) similarly.

Suppose  $\alpha = 0$ . Then

$$\chi_\Omega(\cdot)A_0 \in \mathcal{A}[0, \beta] \subseteq \overline{\mathcal{A}[0, \beta]}^H.$$

Conversely, suppose  $\chi_\Omega(\cdot)A_0 \in \overline{\mathcal{A}[\alpha, \beta]}^H$ . Then there exists a sequence

$$a_k(\cdot) = \chi_{\Omega_k^c}(\cdot)A_0 + \chi_{\Omega_k}(\cdot)A_1 \in \mathcal{A}[\alpha, \beta], \quad \alpha|\Omega| \leq |\Omega_k| \leq \beta|\Omega|, \quad k \geq 1,$$

such that  $a_k(\cdot) \xrightarrow{H} \chi_\Omega(\cdot)A_0$ . We may let

$$\chi_{\Omega_k}(\cdot) \rightarrow g(\cdot), \quad \text{weak* in } L^\infty(\Omega; \mathbb{S}^n),$$

with

$$s \equiv \frac{1}{|\Omega|} \int_\Omega g(x)dx = \lim_{k \rightarrow \infty} \frac{|\Omega_k|}{|\Omega|} \in [\alpha, \beta].$$

Then by (vi) of listed properties of  $H$ -convergence, we have

$$A_0 \leq [1 - g(\cdot)]A_0 + g(\cdot)A_1, \quad A_0^{-1} \leq [1 - g(\cdot)]A_0^{-1} + g(\cdot)A_1^{-1}.$$

Thus, integrating each side, one has

$$A_0 \leq (1 - s)A_0 + sA_1, \quad A_0^{-1} \leq (1 - s)A_0^{-1} + sA_1^{-1}.$$

Since  $A_0 \neq A_1$ , the above hold only if  $s = 0$ . Hence,  $\alpha = 0$ , proving the conclusion.  $\square$

We now formulate the following problem which is called an  $H$ -relaxation of Problem  $(\underline{\Delta}[\alpha, \beta])$ , with  $H$  indicating that the relaxation is in the sense of  $H$ -convergence.

**Problem**  $(\underline{\Delta}^H[\alpha, \beta])$ . Find an  $\underline{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$  such that

$$\lambda_{\underline{a}(\cdot)} = \inf_{a(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H} \lambda_a(\cdot) \quad (4.26)$$

Any  $\underline{a}(\cdot)$  satisfying (4.26) is called an optimal control of Problem  $(\underline{\Delta}^H[\alpha, \beta])$ , which is also called an optimal  $H$ -relaxed control of Problem  $(\underline{\Delta}[\alpha, \beta])$ . The superscript “ $H$ ” indicates the  $H$ -relaxation.

Problem  $(\underline{\Delta}^H[\alpha, \beta])$  is a kind of relaxation for control problems in the coefficients, we refer the readers the following works: [7, 9, 21, 23, 25, 28, 33].

We first have the following existence theorem.

**Theorem 4.4** Problem  $(\underline{\Delta}^H[\alpha, \beta])$  admits an optimal control  $\underline{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ .

**Proof** Let  $a_k(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$  be a minimizing sequence of Problem  $(\underline{\Delta}^H[\alpha, \beta])$  with  $(\lambda_k, y_k(\cdot))$  being the corresponding principal eigen-pair. Thus,

$$\begin{cases} -\nabla \cdot (a_k(x) \nabla y_k(x)) = \lambda_k y_k(x), & x \in \Omega, \\ y_k|_{\partial\Omega} = 0, \end{cases} \quad (4.27)$$

$$y_k(x) \geq 0, \quad x \in \Omega, \quad \int_{\Omega} |y_k(x)|^2 dx = 1, \quad (4.28)$$

and

$$\lim_{k \rightarrow \infty} \lambda_k = \underline{\lambda} \equiv \inf_{a(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H} \lambda_{a(\cdot)}.$$

Since  $y_k(\cdot)$  is uniformly bounded in  $W_0^{1,2}(\Omega)$ , and  $\mathcal{M}[\mu_0, \mu_1]$  is sequentially compact under  $H$ -convergence (Property (i) of  $H$ -convergence listed in Sect. 4.1), we may assume that

$$y_k(\cdot) \rightarrow \underline{y}(\cdot), \quad \text{weakly in } W_0^{1,2}(\Omega), \quad \text{strongly in } L^2(\Omega), \quad (4.29)$$

and

$$a_k(\cdot) \xrightarrow{H} \underline{a}(\cdot), \quad (4.30)$$

for some  $\underline{y}(\cdot) \in W_0^{1,2}(\Omega)$  and  $\underline{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ . Clearly,

$$\underline{y}(x) \geq 0, \quad x \in \Omega, \quad \int_{\Omega} |\underline{y}(x)|^2 dx = 1, \quad (4.31)$$

and

$$\begin{cases} -\nabla \cdot (\underline{a}(x) \nabla \underline{y}(x)) = \underline{\lambda} \underline{y}(x), & x \in \Omega, \\ \underline{y}|_{\partial\Omega} = 0, \end{cases} \quad (4.32)$$

Hence,  $\lambda_{\underline{a}(\cdot)} = \underline{\lambda}$  and  $\underline{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$  is an optimal control.  $\square$

Now, we state the following necessary conditions for an optimal control of Problem  $(\underline{\Delta}^H[\alpha, \beta])$ .

**Theorem 4.5** Let  $\underline{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$  be an optimal control of Problem  $(\underline{\Delta}^H[\alpha, \beta])$  with  $(\lambda, \underline{y}(\cdot)) \in [\mu_0, \mu_1] \times W_0^{1,2}(\Omega)$  being the corresponding principal eigen pair. Then

$$\begin{aligned} \underline{\lambda} &= \int_{\Omega} |\underline{a}(x)^{\frac{1}{2}} \nabla \underline{y}(x)|^2 dx = \sup_{b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H} \int_{\Omega} |b(x)^{-\frac{1}{2}} \underline{a}(x) \nabla \underline{y}(x)|^2 dx \\ &= \sup_{b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H} \int_{\Omega} |b(x)^{-\frac{1}{2}} \underline{a}(x) \nabla \underline{y}(x)|^2 dx, \end{aligned} \quad (4.33)$$

and

$$\int_{\Omega} \langle [\underline{a}(x) - \underline{a}(x)b(x)^{-1}\underline{a}(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx \geq 0, \quad \forall b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H, \quad (4.34)$$

Equivalently,

$$\int_{\Omega} \langle [\underline{a}(x) - \underline{a}(x)b(x)^{-1}\underline{a}(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx \geq 0, \quad \forall b(\cdot) \in \mathcal{A}[\alpha, \beta]. \quad (4.35)$$

When  $\alpha = 0$  and  $\beta = 1$ , the following also holds:

$$\langle \underline{a}(x) \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \geq \langle B^{-1} \underline{a}(x) \nabla \underline{y}(x), \underline{a}(x) \nabla \underline{y}(x) \rangle, \quad \text{a.e.}x \in \Omega, \quad B \in \overline{\{A_0, A_1\}}^H, \quad (4.36)$$

which is equivalent to

$$\langle \underline{a}(x) \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \geq \langle A_i^{-1} \underline{a}(x) \nabla \underline{y}(x), \underline{a}(x) \nabla \underline{y}(x) \rangle, \quad \text{a.e.}x \in \Omega, \quad i = 0, 1. \quad (4.37)$$

This is a kind of maximum principle for the optimal control of Problem  $(\underline{\Lambda}^H[\alpha, \beta])$ . We point out that when  $0 < \alpha$  or  $\beta < 1$ , one could not get (4.36). Also, it is clear that if (4.37) holds, then by  $H$ -convergence,

$$\langle \underline{a}(x) \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \geq \langle b(x)^{-1} \underline{a}(x) \nabla \underline{y}(x), \underline{a}(x) \nabla \underline{y}(x) \rangle, \quad \text{a.e.}x \in \Omega, \quad \forall b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H. \quad (4.38)$$

In other words, in some sense, (4.36), (4.37) and (4.38) are mutually equivalent. To prove this theorem, we need several lemmas.

**Lemma 4.6** *The metric  $\rho(\cdot, \cdot)$  on  $\mathcal{M}[\mu_0, \mu_1]$  defined by (4.8) is uniformly continuous in the following sense: For any  $\varepsilon > 0$ , there exists a  $\delta > 0$  (only depending on  $\varepsilon > 0$ ) such that*

$$\rho(a(\cdot), b(\cdot)) < \varepsilon, \quad \forall a(\cdot), b(\cdot) \in \mathcal{M}[\mu_0, \mu_1], \quad \|a(\cdot) - b(\cdot)\|_1 < \delta. \quad (4.39)$$

Consequently, if  $\{a_k(\cdot)\}$  and  $\{b_k(\cdot)\}$  are two sequences in  $\mathcal{M}[\mu_0, \mu_1]$  such that

$$a_k(\cdot) \xrightarrow{H} a^*(\cdot), \quad \|a_k(\cdot) - b_k(\cdot)\|_1 \rightarrow 0. \quad (4.40)$$

Then  $b_k(\cdot) \xrightarrow{H} a^*(\cdot)$ .

**Proof** Since  $\mathcal{M}[\mu_0, \mu_1]$  is sequentially compact under  $H$ -convergence, it suffices to show that (4.40) implies  $b_k(\cdot) \xrightarrow{H} a^*(\cdot)$ .

Let  $f \in W^{-1,2}(\Omega)$ . Consider

$$\begin{cases} -\nabla \cdot (a_k(x) \nabla y_k(x)) = f, & \text{in } \Omega, \\ y_k|_{\partial\Omega} = 0 \end{cases} \quad (4.41)$$

and

$$\begin{cases} -\nabla \cdot (b_k(x) \nabla z_k(x)) = f, & \text{in } \Omega, \\ z_k|_{\partial\Omega} = 0. \end{cases} \quad (4.42)$$

Then

$$y_k(\cdot) \rightarrow y^*(\cdot), \quad \text{weakly in } W_0^{1,2}(\Omega)$$

with

$$\begin{cases} -\nabla \cdot (a^*(x) \nabla y^*(x)) = f, & \text{in } \Omega, \\ y^*|_{\partial\Omega} = 0. \end{cases}$$

Then, thanks to Theorem 2.3, for some  $p > 2$  and  $C = C_f$ , the following holds:

$$\|\nabla y_k\|_{L^p(\Omega)} \leq C, \quad \|\nabla z_k\|_{L^p(\Omega)} \leq C.$$

We note that (making use of the Dominated Convergence Theorem)

$$\|a_k(\cdot) - b_k(\cdot)\|_{\frac{2p}{p-2}} \rightarrow 0.$$

It holds that

$$\begin{aligned}
0 &= \int_{\Omega} a_k(x) \nabla y_k(x) \cdot \nabla (y_k(x) - z_k(x)) dx - \int_{\Omega} b_k(x) \nabla z_k(x) \cdot \nabla (y_k(x) - z_k(x)) dx \\
&= \int_{\Omega} a_k(x) (\nabla y_k(x) - \nabla z_k(x)) \cdot (\nabla y_k(x) - \nabla z_k(x)) dx \\
&\quad + \int_{\Omega} (a_k(x) - b_k(x)) \nabla z_k(x) \cdot \nabla (y_k(x) - z_k(x)) dx \\
&\geq \mu_0 \|\nabla (y_k(\cdot) - z_k(\cdot))\|_2 - \|a_k(\cdot) - b_k(\cdot)\|_{\frac{2p}{p-2}} \|\nabla (z_k(\cdot))\|_p (\|\nabla (y_k(\cdot) - z_k(\cdot))\|_p).
\end{aligned}$$

Therefore,

$$z_k(\cdot) - y_k(\cdot) \rightarrow 0, \quad \text{strongly in } W_0^{1,2}(\Omega).$$

This implies

$$z_k(\cdot) \rightarrow y^*(\cdot), \quad \text{weakly in } W_0^{1,2}(\Omega).$$

That is,  $b_k(\cdot) \xrightarrow{H} a^*(\cdot)$ . This completes the proof.  $\square$

The above result shows that there exists a non-decreasing function  $h : [0, \infty) \rightarrow [0, \infty)$  with  $h(0) = h(0+) = 0$  such that

$$\rho(a(\cdot), b(\cdot)) \leq h(\|a(\cdot) - b(\cdot)\|_1), \quad \forall a(\cdot), b(\cdot) \in \mathcal{M}[\mu_0, \mu_1]. \quad (4.43)$$

Such a relation will be used below.

#### Lemma 4.7

(i) Let  $\{\Omega_i\}$  be a sequence of measurable subsets of  $\Omega$  such that

$$|\Omega_i| \rightarrow \gamma |\Omega|$$

and

$$a_i(\cdot) = A_0 + \chi_{\Omega_i}(\cdot)(A_1 - A_0) \xrightarrow{H} a^*(\cdot).$$

Then  $a^*(\cdot) \in \overline{\mathcal{A}[\gamma, \gamma]}^H$ .

(ii) For any  $a^*(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ , there is a  $\gamma \in [\alpha, \beta]$  such that  $a^*(\cdot) \in \overline{\mathcal{A}[\gamma, \gamma]}^H$ .

**Proof** (i) For any  $i$ , we can choose a measurable set  $\tilde{\Omega}_i \subseteq \Omega$  such that  $|\tilde{\Omega}_i| = \gamma |\Omega|$  and

$$\int_{\Omega} |\chi_{\Omega_i}(x) - \chi_{\tilde{\Omega}_i}(x)| dx = \left| |\Omega_i| - \gamma |\Omega| \right|.$$

In fact, we can choose  $\tilde{\Omega}_i \subseteq \Omega_i$  if  $|\Omega_i| \geq \gamma |\Omega|$  and  $\tilde{\Omega}_i \supset \Omega_i$  if  $|\Omega_i| < \gamma |\Omega|$ . Let

$$\tilde{a}_i(\cdot) = A_0 + \chi_{\tilde{\Omega}_i}(\cdot)(A_1 - A_0) \in \mathcal{A}[\gamma, \gamma].$$

Then  $\|\tilde{a}_i(\cdot) - a_i(\cdot)\|_1 \rightarrow 0$ . By Lemma 4.6, we get  $\tilde{a}_i(\cdot) \xrightarrow{H} a^*(\cdot)$  and consequently  $a^*(\cdot) \in \overline{\mathcal{A}[\gamma, \gamma]}^H$ .

(ii) The result follows directly from (i).  $\square$

**Lemma 4.8** Let  $\{\Omega_i \subseteq \Omega \mid 1 \leq i \leq m\}$  be a partition of  $\Omega$ , i.e., it is a set of mutually disjoint domain such that

$$|\Omega \setminus \bigcup_{i=1}^m \Omega_i| = 0. \quad (4.44)$$

(i) For  $i = 1, 2, \dots, m$ , let  $a_\varepsilon^i(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$  be such that

$$a_\varepsilon^i(\cdot) \xrightarrow{H} a_0^i(\cdot), \quad 1 \leq i \leq m. \quad (4.45)$$

Then

$$\sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_\varepsilon^i(\cdot) \xrightarrow{H} \sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_0^i(\cdot). \quad (4.46)$$

(ii) For  $i = 1, 2, \dots, m$ , let  $a_i(\cdot)|_{\Omega_i} \in \overline{\mathcal{A}_{\Omega_i}[0, 1]}^H$ . Then

$$\sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_i(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H \quad (4.47)$$

if and only if for some  $\gamma_1, \gamma_2, \dots, \gamma_m \in [0, 1]$ , one has  $a_i(\cdot)|_{\Omega_i} \in \overline{\mathcal{A}_{\Omega_i}[\gamma_i, \gamma_i]}^H$  and

$$\alpha|\Omega| \leq \sum_{i=1}^m \gamma_i |\Omega_i| \leq \beta|\Omega|. \quad (4.48)$$

**Proof** (i) By the locality of  $H$ -convergence (see Property (ii) in §4.1), we get  $a_\varepsilon^i(\cdot) \xrightarrow{H} a_0^i(\cdot)$  on  $\Omega_i$  for every  $i = 1, 2, \dots, m$ .

On the other hand, by the compactness of  $H$ -convergence (see Property (i) in §4.1), along a subsequence,  $\sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_\varepsilon^i(\cdot) \xrightarrow{H} a^*(\cdot)$  on  $\Omega$  for some  $a^*(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ .

Then by locality,  $a_\varepsilon^i(\cdot) \xrightarrow{H} a^*(\cdot)$  on  $\Omega_i$  for every  $i = 1, 2, \dots, m$ . By the uniqueness of  $H$ -convergent limit (see Property (iii) in §4.1),  $a^*(\cdot) = a_0^i(\cdot)$  on  $\Omega_i$  ( $1 \leq i \leq m$ ). Thus,  $a^*(\cdot) = \sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_0^i(\cdot)$  on  $\Omega$ . Consequently, (4.46) holds, not only in the sense of subsequence.

(ii) **Sufficiency.** For  $i = 1, 2, \dots, m$ , we have  $\Omega_i^k \subset \Omega_i$  such that  $|\Omega_i^k| = \gamma_i |\Omega_i|$  and

$$a_i^k(\cdot) = A_0 + \chi_{\Omega_i^k}(\cdot)(A_1 - A_0) \xrightarrow{H} a_i(\cdot), \quad \text{on } \Omega_i.$$

Thus, by (i),

$$\sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_i^k(\cdot) \xrightarrow{H} \sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_i(\cdot).$$

On the other hand,

$$\frac{1}{|\Omega|} \left| \bigcup_{i=1}^m \Omega_i^k \right| = \frac{1}{|\Omega|} \sum_{i=1}^m \gamma_i |\Omega_i| \equiv \gamma \in [\alpha, \beta].$$

Therefore,  $\sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_i(\cdot) \in \overline{\mathcal{A}[\gamma, \gamma]}^H \subseteq \overline{\mathcal{A}[\alpha, \beta]}^H$ .

**Necessity.** By Lemma 4.7, there is  $\gamma \in [\alpha, \beta]$  such that  $\sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_i(\cdot) \in \overline{\mathcal{A}[\gamma, \gamma]}^H$ . Thus, there is a sequence  $\{E_k\}$  of measurable subsets of  $\Omega$  such that  $|E_k| = \gamma|\Omega|$  and

$$a_k(\cdot) = A_0 + \chi_{E_k}(\cdot)(A_1 - A_0) \xrightarrow{H} \sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_i(\cdot).$$

Then, by locality,  $a_k(\cdot) \xrightarrow{H} a_i(\cdot)$  on  $\Omega_i$ . On the other hand, we can suppose that for every  $i = 1, 2, \dots, m$ ,  $|E_k \cap \Omega_i|$  converges to  $\gamma_i|\Omega_i|$  for some  $\gamma_i \in [0, 1]$ . Thus  $a_i(\cdot)|_{\Omega_i} \in \overline{\mathcal{A}_{\Omega_i}[\gamma_i, \gamma_i]}^H$  and

$$\sum_{i=1}^m \gamma_i |\Omega_i| = \lim_{k \rightarrow +\infty} \sum_{i=1}^m |E_k \cap \Omega_i| = \lim_{k \rightarrow +\infty} |E_k| = \gamma |\Omega|.$$

We get the proof.  $\square$

**Remark 4.9** We would like to mention that  $\gamma$  in Lemma 4.7(ii) might be not unique. Therefore, it is possible that although  $\sum_{i=1}^m \chi_{\Omega_i}(\cdot) a_i(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$  and  $a_i(\cdot)|_{\Omega_i} \in \overline{\mathcal{A}_{\Omega_i}[\gamma_i, \gamma_i]}^H$ , but  $\frac{1}{|\Omega|} \sum_{i=1}^m \gamma_i |\Omega_i| \notin [\alpha, \beta]$ .

The following result is an extension of Theorem 4.2 (i), replacing  $A$  and  $B$  by  $a(\cdot)$  and  $b(\cdot)$ , respectively.

**Lemma 4.10** *Let  $a(\cdot), b(\cdot) \in \mathcal{M}[\mu_0, \mu_1]$ ,  $\theta \in [0, 1]$  and  $e \in S^{n-1}$ . For any small  $\varepsilon > 0$ , define*

$$\mathcal{H}_\varepsilon[a(\cdot), b(\cdot); \theta, e](x) = \begin{cases} a(x), & \left\{ \frac{\langle x, e \rangle}{\varepsilon} \right\} \in [\theta, 1), \\ b(x), & \left\{ \frac{\langle x, e \rangle}{\varepsilon} \right\} \in [0, \theta). \end{cases} \quad (4.49)$$

Then as  $\varepsilon \rightarrow 0$ ,

$$\begin{aligned} \mathcal{H}_\varepsilon[a(\cdot), b(\cdot); \theta, e](\cdot) &\xrightarrow{H} \mathcal{H}[a(\cdot), b(\cdot); \theta, e](\cdot) \\ &\equiv a(\cdot) - \theta[a(\cdot) - b(\cdot)] - \frac{\theta(1-\theta)[a(\cdot) - b(\cdot)]ee^T[a(\cdot) - b(\cdot)]}{e^T\{b(\cdot) + \theta[a(\cdot) - b(\cdot)]\}e}. \end{aligned} \quad (4.50)$$

The proof of the above lemma essentially follows from that of [24], Proposition 2.1. Based on the above, we further have the following result.

**Lemma 4.11** *Let  $\gamma_i \in [0, 1]$ ,  $a_i(\cdot) \in \overline{\mathcal{A}[\gamma_i, \gamma_i]}^H$  ( $i = 1, 2$ ). Then for any  $\theta \in [0, 1]$  and  $e \in S^{n-1}$ ,*

$$b(\cdot) \equiv a_1(\cdot) - \theta[a_1(\cdot) - a_2(\cdot)] - \frac{\theta(1-\theta)[a_1(\cdot) - a_2(\cdot)]ee^T[a_1(\cdot) - a_2(\cdot)]}{e^T\{a_2(\cdot) + \theta[a_1(\cdot) - a_2(\cdot)]\}e} \in \overline{\mathcal{A}[\gamma, \gamma]}^H, \quad (4.51)$$

where  $\gamma = (1-\theta)\gamma_1 + \theta\gamma_2$ . Consequently,

$$\mathcal{H}\left[\overline{\mathcal{A}[\alpha, \beta]}^H, \overline{\mathcal{A}[\alpha, \beta]}^H; [0, 1], S^{n-1}\right] \subseteq \overline{\mathcal{A}[\alpha, \beta]}^H. \quad (4.52)$$

**Proof** We need only to consider the case  $\theta \in (0, 1)$ , which is fixed below. Take an  $e \in S^{n-1}$ . For simplicity of notation, denote  $\theta_1 = 1 - \theta$ ,  $\theta_2 = \theta$ . For any  $k \geq 1$ , denote

$$\mathcal{Q}_k^1 = \{x \in \Omega \mid \{k \langle x, e \rangle\} \in [\theta, 1)\}, \quad \mathcal{Q}_k^2 = \{x \in \Omega \mid \{k \langle x, e \rangle\} \in [0, \theta)\} \quad (4.53)$$

and define

$$b_k(\cdot) = \chi_{\mathcal{Q}_k^1}(\cdot)a_1(\cdot) + \chi_{\mathcal{Q}_k^2}(\cdot)a_2(\cdot). \quad (4.54)$$

Then

$$\chi_{\mathcal{Q}_k^i}(\cdot) \rightarrow \theta_i \chi_{\Omega}(\cdot), \quad \text{weakly in } L^2(\Omega), \quad i = 1, 2. \quad (4.55)$$

Moreover, by Lemma 4.10,  $b_k(\cdot) \xrightarrow{H} b(\cdot)$ .

On the other hand, for  $i = 1, 2$ , there is a sequence  $\{E_j^i\}_{j \geq 1}$  of measurable subsets of  $\Omega$  such that  $|E_j^i| = \gamma_i |\Omega|$ ,

$$\chi_{E_j^i}(\cdot) \rightarrow \sigma_i(\cdot), \quad \text{weakly in } L^2(\Omega), \quad (4.56)$$

and

$$a_j^i(\cdot) = A_0 + \chi_{E_j^i}(\cdot)(A_1 - A_0) \xrightarrow{H} a_i(\cdot). \quad (4.57)$$

By (4.55),

$$\sum_{i=1}^2 \int_{\mathcal{Q}_k^i} \sigma_i(x) dx \rightarrow \sum_{i=1}^2 \int_{\Omega} \theta_i \sigma_i(x) dx = \gamma |\Omega|, \quad (4.58)$$

Thus, for  $m \geq 1$ , we have  $k_m \geq 1$  such that (c.f. (4.8))

$$\rho(b_{k_m}(\cdot), b(\cdot)) \leq \frac{1}{m} \quad (4.59)$$

and

$$\left| \sum_{i=1}^2 \int_{\mathcal{Q}_{k_m}^i} \sigma_i(x) dx - \gamma |\Omega| \right| \leq \frac{1}{m}. \quad (4.60)$$

By Lemma 4.7, as  $j \rightarrow \infty$ ,

$$\sum_{i=1}^2 \chi_{\mathcal{Q}_{k_m}^i}(\cdot) a_j^i(\cdot) \xrightarrow{H} \sum_{i=1}^2 \chi_{\mathcal{Q}_{k_m}^i}(\cdot) a_i(\cdot) = b_{k_m}(\cdot). \quad (4.61)$$

Then, by (4.61) and (4.56), we have  $j_m \geq 1$  such that

$$\rho \left( \sum_{i=1}^2 \chi_{\mathcal{Q}_{k_m}^i}(\cdot) a_{j_m}^i(\cdot), b_{k_m}(\cdot) \right) \leq \frac{1}{m} \quad (4.62)$$

and

$$\sum_{i=1}^2 \left| \int_{\mathcal{Q}_{k_m}^i} (\chi_{E_{j_m}^i}(x) - \sigma_i(x)) dx \right| \leq \frac{1}{m}. \quad (4.63)$$

Denote

$$B_m(\cdot) = \sum_{i=1}^2 \chi_{\mathcal{Q}_{k_m}^i}(\cdot) a_{j_m}^i(\cdot).$$

Then

$$B_m(\cdot) = A_0 + \chi_{\Omega_m}(\cdot)(A_1 - A_0), \quad \Omega_m = \bigcup_{i=1}^2 (Q_{k_m}^i \cap E_{j_m}^i).$$

By (4.59) and (4.62),  $B_m(\cdot) \xrightarrow{H} b(\cdot)$ . By (4.61) and (4.63),  $\Omega_m \rightarrow \gamma|\Omega|$ . Thus, it follows from Lemma 4.7(i) that  $b(\cdot) \in \overline{\mathcal{A}[\gamma, \gamma]}^H$ , proving our claim.  $\square$

We further extend the above result to the following (replacing  $e$  by  $\xi(\cdot)$ ).

**Lemma 4.12** *Assume  $a(\cdot), b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ . Let  $\Omega_1, \Omega_2, \dots, \Omega_m$  be a partition of  $\Omega$  and*

$$\xi(x) = \sum_{i=1}^m \chi_{\Omega_i}(x) \xi_i,$$

with  $\xi_i \in S^{n-1}$  ( $1 \leq i \leq m$ ). Then for any  $\theta \in [0, 1]$ ,

$$a(\cdot) - \theta[a(\cdot) - b(\cdot)] - \frac{\theta(1-\theta)(a(\cdot) - b(\cdot))\xi(\cdot)\xi(\cdot)^T(a(\cdot) - b(\cdot))}{\xi(\cdot)^T(b(\cdot) + \theta[a(\cdot) - b(\cdot)])\xi(\cdot)} \in \overline{\mathcal{A}[\alpha, \beta]}^H. \quad (4.64)$$

**Proof** For  $m = 1, 2, \dots, m$ , denote

$$b_i(\cdot) = a(\cdot) - \theta[a(\cdot) - b(\cdot)] - \frac{\theta(1-\theta)[a(\cdot) - b(\cdot)]\xi_i \xi_i^T[a(\cdot) - b(\cdot)]}{\xi_i^T[b(\cdot) + \theta[a(\cdot) - b(\cdot)]]\xi_i}.$$

Since  $a(\cdot), b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ , by Lemma 4.8, for  $j = 1, 2$ , there exist  $\gamma_{j1}, \gamma_{j2}, \dots, \gamma_{jm} \in [0, 1]$  such that  $a_j(\cdot)|_{\Omega_i} \in \overline{\mathcal{A}_{\Omega_i}[\gamma_{ji}, \gamma_{ji}]}^H$ , and

$$\alpha|\Omega| \leq \sum_{i=1}^m \gamma_{ji} |\Omega_i| \leq \beta|\Omega|, \quad (4.65)$$

where we denote  $a_1(\cdot) = a(\cdot)$ ,  $a_2(\cdot) = b(\cdot)$  for notation simplicity. By Lemma 4.11,  $b_i(\cdot)|_{\Omega_i} \in \overline{\mathcal{A}_{\Omega_i}[\gamma_i, \gamma_i]}^H$  with  $\gamma_i = (1-\theta)\gamma_{1i} + \theta\gamma_{2i}$  ( $i = 1, 2, \dots, m$ ). Since 5

$$\alpha|\Omega| \leq \sum_{i=1}^m \gamma_i |\Omega_i| = (1-\theta) \sum_{i=1}^m \gamma_{1i} |\Omega_i| + \theta \sum_{i=1}^m \gamma_{2i} |\Omega_i| \leq \beta|\Omega|,$$

we get from Lemma 4.8 that

$$\sum_{i=1}^m \chi_{\Omega_i}(\cdot) b_i(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H.$$

That is, (4.64) holds.  $\square$

Now, we are ready to prove Theorem 4.5.

**Proof of Theorem 4.5** Fix  $b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ . Let  $\Omega_1, \Omega_2, \dots, \Omega_m$  be a partition of  $\Omega$  and

$$\xi(x) = \sum_{k=1}^m \chi_{\Omega_k}(x) \xi_k, \quad (4.66)$$

with  $\xi_k \in S^{n-1}$  ( $1 \leq k \leq m$ ). Then, by Lemma 4.11, for any  $\theta \in (0, 1)$ ,

$$a_\theta(\cdot) \equiv \underline{a}(\cdot) - \theta[\underline{a}(\cdot) - b(\cdot)] - \frac{\theta(1-\theta)[\underline{a}(\cdot) - b(\cdot)]\xi(\cdot)\xi(\cdot)^T[\underline{a}(\cdot) - b(\cdot)]}{\xi(\cdot)^T(b(\cdot) + \theta[\underline{a}(\cdot) - b(\cdot)])\xi(\cdot)} \in \overline{\mathcal{A}[\alpha, \beta]}^H. \quad (4.67)$$

By the optimality of  $\underline{a}(\cdot)$ , we have

$$\begin{aligned} \int_{\Omega} \langle \underline{a}(x) \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx &= \lambda_{\underline{a}(\cdot)} \leq \lambda_{a_\theta(\cdot)} \\ &= \inf_{\substack{y(\cdot) \in W_0^{1,2}(\Omega) \\ \|y(\cdot)\|_2=1}} \int_{\Omega} \langle a_\theta(x) \nabla y(x), \nabla y(x) \rangle dx \leq \int_{\Omega} \langle a_\theta(x) \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx. \end{aligned}$$

Therefore, by the minimality of  $\underline{a}(\cdot)$ , one has

$$\begin{aligned} 0 &\leq \lim_{\theta \rightarrow 0^+} \int_{\Omega} \left\langle \frac{a_\theta(x) - \underline{a}(x)}{\theta} \nabla \underline{y}(x), \nabla \underline{y}(x) \right\rangle dx \\ &= - \int_{\Omega} \left\langle \left( \underline{a}(x) - b(x) + \frac{(\underline{a}(x) - b(x))\xi(x)\xi(x)^T(\underline{a}(x) - b(x))}{\xi(x)^T b(x) \xi(\cdot)} \right) \nabla \underline{y}(x), \nabla \underline{y}(x) \right\rangle dx. \end{aligned}$$

That is,

$$\begin{aligned} \int_{\Omega} \left\langle [b(x) - \underline{a}(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \right\rangle dx &\geq \int_{\Omega} \frac{|\xi(x)^T [\underline{a}(x) - b(x)] \nabla \underline{y}(x)|^2}{|b(x)^{\frac{1}{2}} \xi(x)|^2} dx \\ &\geq \int_{\Omega} \left\langle b(x)^{-\frac{1}{2}} [\underline{a}(x) - b(x)] \nabla \underline{y}(x), \frac{b(x)^{\frac{1}{2}} \xi(x)}{|b(x)^{\frac{1}{2}} \xi(x)|} \right\rangle^2 dx. \end{aligned}$$

The above is true for any  $\xi(\cdot)$  of form (4.66). Then by approximation, we obtain

$$\begin{aligned} \int_{\Omega} \langle [b(x) - \underline{a}(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx &\geq \int_{\Omega} |b(x)^{-\frac{1}{2}} (\underline{a}(x) - b(x)) \nabla \underline{y}(x)|^2 dx \\ &= \int_{\Omega} \langle [\underline{a}(x) - b(x)] b(x)^{-1} [\underline{a}(x) - b(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx \\ &= \int_{\Omega} \langle [\underline{a}(x) b(x)^{-1} \underline{a}(x) - 2\underline{a}(x) + b(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx. \end{aligned}$$

Therefore, (4.34) holds. Consequently, we have (4.35).

Now, we show that (4.35) also implies (4.34). Suppose that (4.35) holds. For any  $b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ , there is a sequence  $\Omega_k \subseteq \Omega$  with

$$\alpha|\Omega| \leq |\Omega_k| \leq \beta|\Omega|, \quad k \geq 1,$$

such that as  $k \rightarrow +\infty$ ,

$$b_k(\cdot) \equiv \chi_{\Omega_k^c}(\cdot) A_0 + \chi_{\Omega_k}(\cdot) A_1 \xrightarrow{H} b(\cdot). \quad (4.68)$$

We can assume that

$$\chi_{\Omega_k}(\cdot) \rightarrow g(\cdot), \quad \text{weakly in } L^2(\Omega).$$

Then

$$b_k(\cdot)^{-1} = \chi_{\Omega_k^c}(\cdot) A_0^{-1} + \chi_{\Omega_k}(\cdot) A_1^{-1} \rightarrow [1 - g(\cdot)] A_0^{-1} + g(\cdot) A_1^{-1}, \quad \text{weakly in } L^2(\Omega).$$

From (4.10), we get that

$$b(\cdot)^{-1} \leq [1 - g(\cdot)]A_0^{-1} + g(\cdot)A_1^{-1}. \quad (4.69)$$

Therefore,

$$\begin{aligned} & \int_{\Omega} \langle \underline{a}(x)b(x)^{-1}\underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx \\ & \leq \int_{\Omega} \langle \underline{a}(x)[(1 - g(x))A_0^{-1} + g(x)A_1^{-1}]\underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx \\ & = \lim_{k \rightarrow \infty} \int_{\Omega} \langle \underline{a}(x)[\chi_{\Omega_k^c}(x)A_0^{-1} + \chi_{\Omega_k}(x)A_1^{-1}]\underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx \\ & \leq \int_{\Omega} \langle \underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx, \end{aligned}$$

where the last inequality follows from (4.35). Hence, we get (4.34). Then it follows that

$$\begin{aligned} & \sup_{b(\cdot) \in \mathcal{A}[\alpha, \beta]} \int_{\Omega} \langle \underline{a}(x)b(x)^{-1}\underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx \\ & = \max_{b(\cdot) \in \mathcal{A}[\alpha, \beta]^H} \int_{\Omega} \langle \underline{a}(x)b(x)^{-1}\underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx \\ & = \int_{\Omega} \langle \underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx = \underline{\lambda}. \end{aligned} \quad (4.70)$$

Hence, (4.33) holds.

Next, in the case that  $\alpha = 0$  and  $\beta = 1$ , for any sub-domain  $\Omega_{\varepsilon} \subseteq \Omega$  with  $|\Omega_{\varepsilon}| = \varepsilon$ , and  $b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ , let (noting Lemma 4.8 (ii))

$$b_{\varepsilon}(\cdot) = \chi_{\Omega_{\varepsilon}}(\cdot)b(\cdot) + \chi_{\Omega_{\varepsilon}^c}(\cdot)\underline{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H.$$

Taking such a  $b_{\varepsilon}(\cdot)$  in the above, we obtain

$$\int_{\Omega_{\varepsilon}} \langle \underline{a}(x)b(x)^{-1}\underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx \leq \int_{\Omega_{\varepsilon}} \langle \underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \, dx.$$

Then, using Lebesgue's density theorem, we obtain (4.38). In particular, (4.37) holds. Moreover, similar to that (4.34) and (4.35) are equivalent, (4.37) is equivalent to (4.36).  $\square$

#### 4.4 Optimality system

Let us now take a closer look at (4.34)/(4.70). Note that any  $b(\cdot) \in \mathcal{A}[\alpha, \beta]$  has the following form:

$$b(\cdot) = \chi_{\Omega_1^c}(\cdot)A_0 + \chi_{\Omega_1}(\cdot)A_1,$$

for some  $\Omega_1 \subseteq \Omega$ , with  $\alpha|\Omega| \leq |\Omega_1| \leq \beta|\Omega|$ . Then there is a sequence of  $\Omega_k \subseteq \Omega$  such that  $\alpha|\Omega| \leq |\Omega_k| \leq \beta|\Omega|$ ,

$$\chi_{\Omega_k^c}(\cdot)A_0 + \chi_{\Omega_k}(\cdot)A_1 \xrightarrow{H} \underline{a}(\cdot)$$

and

$$\chi_{\Omega_j}(\cdot) \rightarrow \underline{a}(\cdot), \quad \text{weakly in } L^2(\Omega; [0, 1]).$$

Then  $\underline{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$  (see (3.1) for the definition) and by property (vi) in Sect. 4.1,

$$\underline{a}(\cdot)^{-1} \leq A_0^{-1} + \underline{\sigma}(\cdot)(A_1^{-1} - A_0^{-1}), \quad x \in \Omega.$$

Consequently, (4.70) is equivalent to the following:

$$\begin{aligned} & \int_{\Omega} \left\langle (A_0^{-1} + \underline{\sigma}(x)(A_1^{-1} - A_0^{-1}))\underline{a}(x)\nabla \underline{y}(x), \underline{a}(x)\nabla \underline{y}(x) \right\rangle dx \geq \int_{\Omega} \langle \underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx \\ &= \sup_{\alpha|\Omega| \leq |\Omega_1| \leq \beta|\Omega|} \int_{\Omega} \langle \underline{a}(x)[A_0^{-1}\chi_{\Omega_1^c}(x) + A_1^{-1}\chi_{\Omega_1}(x)]\underline{a}(x)\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle dx \\ &= \sup_{\sigma(\cdot) \in \Sigma[\alpha, \beta]} \int_{\Omega} \left\langle (A_0^{-1} + \sigma(x)(A_1^{-1} - A_0^{-1}))\underline{a}(x)\nabla \underline{y}(x), \underline{a}(x)\nabla \underline{y}(x) \right\rangle dx. \end{aligned}$$

Therefore,

$$\begin{aligned} & \int_{\Omega} \underline{\sigma}(x) \left\langle (A_1^{-1} - A_0^{-1})\underline{a}(x)\nabla \underline{y}(x), \underline{a}(x)\nabla \underline{y}(x) \right\rangle dx \\ &= \sup_{\sigma(\cdot) \in \Sigma[\alpha, \beta]} \int_{\Omega} \sigma(x) \left\langle (A_1^{-1} - A_0^{-1})\underline{a}(x)\nabla \underline{y}(x), \underline{a}(x)\nabla \underline{y}(x) \right\rangle dx. \end{aligned}$$

By denoting

$$h(x) = \left\langle (A_0^{-1} - A_1^{-1})\underline{a}(x)\nabla \underline{y}(x), \underline{a}(x)\nabla \underline{y}(x) \right\rangle,$$

one sees that the above becomes

$$\int_{\Omega} \underline{\sigma}(x)h(x)dx = \sup_{\sigma(\cdot) \in \Sigma[\alpha, \beta]} \int_{\Omega} \sigma(x)h(x)dx. \quad (4.71)$$

Thus,  $\underline{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$  solves a maximization problem. For this problem, we have the following proposition.

**Proposition 4.13** *Let  $\underline{\sigma}(\cdot) \in \Sigma[\alpha, \beta]$  satisfy (4.71). Then there are two constants  $\Psi$  and  $\mu_0$  such that*

$$\begin{aligned} \mu_0 &\leq 0, \quad \mu_0^2 + \Psi^2 = 1, \\ \left( \int_{\Omega} \underline{\sigma}(x)dx - t \right) \Psi &\leq 0, \quad \alpha|\Omega| \leq t \leq \beta|\Omega|. \end{aligned}$$

and

$$(\mu_0 h(x) + \Psi) \underline{\sigma}(x) = \max_{0 \leq \theta \leq 1} (\mu_0 h(x) + \Psi)\theta, \quad \text{a.e. } x \in \Omega.$$

**Proof** For any  $\varepsilon > 0$ , define

$$F_{\varepsilon}(\sigma(\cdot)) = \left\{ \left[ \left( \int_{\Omega} [\sigma(x) - \underline{\sigma}(x)]h(x)dx + \varepsilon \right)^+ \right]^2 + \phi \left( \int_{\Omega} \sigma(x)ds \right) \right\}^{\frac{1}{2}},$$

$$\forall \sigma(\cdot) \in \widehat{\Sigma} = \{ \sigma : \Omega \rightarrow [0, 1] \mid \sigma(\cdot) \text{ is measurable} \},$$

where

$$\phi(s) = \min_{\alpha|\Omega| \leq t \leq \beta|\Omega|} |s - t|^2, \quad s \in \mathbb{R}.$$

We have  $|\phi'(s)| = 2\sqrt{\phi(s)}$  and

$$(t - s)\phi'(s) \leq 0, \quad \alpha|\Omega| \leq t \leq \beta|\Omega|.$$

Clearly,

$$F_\varepsilon(\cdot) > 0, \quad \forall \sigma(\cdot) \in \widehat{\Sigma}, \quad F_\varepsilon(\underline{\sigma}(\cdot)) = \varepsilon.$$

By Ekeland's variational principle ([22]), there exists a  $\sigma_\varepsilon(\cdot) \in \widehat{\Sigma}$  such that

$$\begin{aligned} \|\sigma_\varepsilon(\cdot) - \underline{\sigma}(\cdot)\|_2 &\leq \sqrt{\varepsilon}, \\ F_\varepsilon(\sigma(\cdot)) + \sqrt{\varepsilon} \|\sigma(\cdot) - \sigma_\varepsilon(\cdot)\|_2 &\geq F_\varepsilon(\sigma_\varepsilon(\cdot)), \quad \forall \sigma(\cdot) \in \widehat{\Sigma}. \end{aligned}$$

Hence, for any  $\delta \in (0, 1)$  and  $\sigma(\cdot) \in \widehat{\Sigma}$ , one has

$$\sigma_\varepsilon^\delta(\cdot) \equiv \sigma_\varepsilon(\cdot) + \delta[\sigma(\cdot) - \sigma_\varepsilon(\cdot)] \in \widehat{\Sigma},$$

and thus

$$\begin{aligned} -\sqrt{\varepsilon} \|\sigma(\cdot)\|_2 &\leq \frac{F_\varepsilon(\sigma_\varepsilon^\delta(\cdot)) - F_\varepsilon(\sigma_\varepsilon(\cdot))}{\delta} = \frac{F_\varepsilon(\sigma_\varepsilon^\delta(\cdot))^2 - F_\varepsilon(\sigma_\varepsilon(\cdot))^2}{\delta[F_\varepsilon(\sigma_\varepsilon^\delta(\cdot)) + F_\varepsilon(\sigma_\varepsilon(\cdot))]} \\ &= \frac{1}{[F_\varepsilon(\sigma_\varepsilon^\delta(\cdot)) + F_\varepsilon(\sigma_\varepsilon(\cdot))] \delta} \left\{ \left[ \left( \int_{\Omega} [\sigma_\varepsilon^\delta(x) - \underline{\sigma}(x)] h(x) dx + \varepsilon \right)^+ \right]^2 \right. \\ &\quad \left. - \left[ \left( \int_{\Omega} [\sigma_\varepsilon(x) - \underline{\sigma}(x)] h(x) dx + \varepsilon \right)^+ \right]^2 \right. \\ &\quad \left. + \phi \left( \int_{\Omega} \sigma_\varepsilon^\delta(x) dx \right) - \phi \left( \int_{\Omega} \sigma_\varepsilon(x) dx \right) \right\} \\ &\rightarrow \frac{1}{F_\varepsilon(\sigma_\varepsilon(\cdot))} \left[ \left( \int_{\Omega} [\sigma_\varepsilon(x) - \underline{\sigma}(x)] h(x) dx + \varepsilon \right)^+ \int_{\Omega} [\sigma(x) - \sigma_\varepsilon(x)] h(x) dx \right. \\ &\quad \left. + \frac{1}{2} \phi' \left( \int_{\Omega} \sigma_\varepsilon(x) dx \right) \int_{\Omega} [\sigma(x) - \sigma_\varepsilon(x)] dx \right. \\ &\equiv -\mu_0^\varepsilon \int_{\Omega} [\sigma(x) - \sigma_\varepsilon(x)] h(x) dx - \Psi^\varepsilon \int_{\Omega} [\sigma(x) - \sigma_\varepsilon(x)] dx, \end{aligned}$$

with

$$(\mu_0^\varepsilon)^2 + (\Psi^\varepsilon)^2 = 1, \quad \mu_0^\varepsilon \leq 0,$$

$$\left( \int_{\Omega} \sigma_\varepsilon(x) dx - t \right) \Psi^\varepsilon \leq 0, \quad \alpha |\Omega| \leq t \leq \beta |\Omega|.$$

Then along a subsequence, still denoted it by itself, we may let

$$(\mu_0^\varepsilon, \Psi^\varepsilon) \rightarrow (\mu_0, \Psi), \quad \mu_0^2 + \Psi^2 = 1, \quad \mu_0 \leq 0,$$

$$\left( \int_{\Omega} \underline{\sigma}(x) dx - t \right) \Psi \leq 0, \quad \alpha |\Omega| \leq t \leq \beta |\Omega|,$$

and

$$\int_{\Omega} (\sigma(x) - \underline{\sigma}(x)) (\mu_0 h(x) + \Psi) dx \leq 0.$$

Hence, a standard argument applies to get

$$(\mu_0 h(x) + \Psi) \underline{\sigma}(x) = \max_{0 \leq \theta \leq 1} (\mu_0 h(x) + \Psi) \theta, \quad \text{a.e. } x \in \Omega.$$

We obtain our conclusions.  $\square$

Now, we use the above result to make some further analysis on the optimal control  $\underline{\sigma}(\cdot)$  of Problem  $(\underline{\Lambda}^H[\alpha, \beta])$ .

If  $\mu_0 = 0$ , then  $\Psi \neq 0$  and we have  $\underline{\sigma}(\cdot) \equiv 0$  or  $\underline{\sigma}(\cdot) \equiv 1$ . That is  $\underline{a}(x) \equiv A_0$  or  $\underline{a}(x) \equiv A_1$ .

If  $\mu_0 \neq 0$ , then we can suppose  $\mu_0 = -1$  without loss of generality. Thus

$$\underline{\sigma}(x) = \begin{cases} 1, & \text{a.e. } x \in (h(\cdot) < \Psi), \\ 0, & \text{a.e. } x \in (h(\cdot) > \Psi). \end{cases}$$

This implies that

$$|(h(\cdot) \leq \Psi)| \geq \int_{\Omega} \underline{\sigma}(x) dx \geq \alpha |\Omega|,$$

and

$$|(h(\cdot) < \Psi)| \leq \int_{\Omega} \underline{\sigma}(x) dx \leq \beta |\Omega|.$$

Moreover, (when  $\alpha < \beta$ ) we can see that

$$\begin{aligned} \int_{\Omega} \underline{\sigma}(x) dx > \alpha |\Omega| &\Rightarrow \Psi \leq 0; \\ \int_{\Omega} \underline{\sigma}(x) dx < \beta |\Omega| &\Rightarrow \Psi \geq 0; \\ \Psi > 0 &\Rightarrow \int_{\Omega} \underline{\sigma}(x) dx = \alpha |\Omega|; \\ \Psi < 0 &\Rightarrow \int_{\Omega} \underline{\sigma}(x) dx = \beta |\Omega|. \end{aligned}$$

For any  $x \in (h(\cdot) = \Psi)$ , though it is possible that  $\underline{\sigma}(x)$  be any value of  $[0, 1]$ , there are still some information could be used to determine  $\underline{\sigma}(x)$ . For example, if  $|(h(\cdot) \geq \Psi)| = \alpha |\Omega|$ , then  $\underline{\sigma}(x) = 0$  a.e.  $(h(\cdot) \geq \Psi)$ .

On the other hand, when  $\alpha |\Omega| < \int_{\Omega} \underline{\sigma}(x) dx < \beta |\Omega|$  (it will be the case if  $\alpha = 0, \beta = 1$  and neither  $\chi_{\Omega}(\cdot) A_0$  nor  $\chi_{\Omega}(\cdot) A_1$  is optimal), it should hold that  $\Psi = 0$ . At this moment, on the set  $(h(\cdot) = \Psi) \equiv (h(\cdot) = 0)$ ,

$$\left\langle A_1^{-1} \underline{a}(x) \nabla \underline{y}(x), \underline{a}(x) \nabla \underline{y}(x) \right\rangle = \left\langle A_0^{-1} \underline{a}(x) \nabla \underline{y}(x), \underline{a}(x) \nabla \underline{y}(x) \right\rangle. \quad (4.72)$$

Hence, when  $h(x) = 0$ , one has

$$\begin{aligned} \left\langle A_1 \nabla \underline{y}(x), \nabla \underline{y}(x) \right\rangle &\neq \left\langle A_0^{-1} A_1 \nabla \underline{y}(x), A_1 \nabla \underline{y}(x) \right\rangle \Rightarrow \underline{\sigma}(x) \neq 1; \\ \left\langle A_0 \nabla \underline{y}(x), \nabla \underline{y}(x) \right\rangle &= \left\langle A_1^{-1} A_0 \nabla \underline{y}(x), A_0 \nabla \underline{y}(x) \right\rangle \Rightarrow \underline{\sigma}(x) \neq 0. \end{aligned}$$

## 4.5 Maximization problem

Similar to Problem  $(\underline{\Lambda}^H[\alpha, \beta])$ , it is natural to pose the following  $H$ -relaxation of Problem  $(\bar{\Lambda}[\alpha, \beta])$ .

**Problem**  $(\bar{\Lambda}^H[\alpha, \beta])$ . Find an  $\bar{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$  such that

$$\lambda_{\bar{a}(\cdot)} = \sup_{a(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H} \lambda_{a(\cdot)} \quad (4.73)$$

Any  $\bar{a}(\cdot)$  satisfying (4.73) is called an optimal control of Problem  $(\bar{\Lambda}^H[\alpha, \beta])$ , which is also called an optimal  $H$ -relaxed control of Problem  $(\bar{\Lambda}[\alpha, \beta])$ . Due to the properties of  $H$ -convergence, it is not hard to see that there are results for Problem  $(\bar{\Lambda}^H[\alpha, \beta])$  parallel to the minimization problem. Let us state them here.

### Theorem 4.14

- (i) *Problem  $(\bar{\Lambda}^H[\alpha, \beta])$  admits an optimal control  $\bar{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$ .*
- (ii) *Let  $\bar{a}(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H$  be an optimal control of Problem  $(\bar{\Lambda}^H[\alpha, \beta])$  with  $\bar{y}(\cdot) \in W_0^{1,2}(\Omega)$  being the corresponding normalized principal eigenfunction. Then*

$$\begin{aligned}\bar{\lambda} &= \int_{\Omega} |\bar{a}(x)^{\frac{1}{2}} \nabla \bar{y}(x)|^2 dx = \inf_{b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H} \int_{\Omega} |b(x)^{-\frac{1}{2}} \bar{a}(x) \nabla \bar{y}(x)|^2 dx \\ &= \inf_{b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}} \int_{\Omega} |b(x)^{-\frac{1}{2}} \bar{a}(x) \nabla \bar{y}(x)|^2 dx,\end{aligned}\quad (4.74)$$

and

$$\int_{\Omega} \left\langle [\bar{a}(x) - \bar{a}(x)b(x)^{-1}\bar{a}(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \right\rangle dx \leq 0, \quad \forall b(\cdot) \in \overline{\mathcal{A}[\alpha, \beta]}^H. \quad (4.75)$$

In particular,

$$\int_{\Omega} \left\langle [\bar{a}(x) - \bar{a}(x)b(x)^{-1}\bar{a}(x)] \nabla \underline{y}(x), \nabla \underline{y}(x) \right\rangle dx \leq 0, \quad \forall b(\cdot) \in \mathcal{A}[\alpha, \beta]. \quad (4.76)$$

When  $\alpha = 0$  and  $\beta = 1$ , it holds that:

$$\langle \bar{a}(x) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \leq \langle B^{-1} \bar{a}(x) \nabla \bar{y}(x), \bar{a}(x) \nabla \bar{y}(x) \rangle, \quad \text{a.e. } x \in \Omega, \quad B \in \overline{\{A_0, A_1\}}^H. \quad (4.77)$$

In particular,

$$\langle \bar{a}(x) \nabla \bar{y}(x), \nabla \bar{y}(x) \rangle \leq \langle A_i^{-1} \bar{a}(x) \nabla \bar{y}(x), \bar{a}(x) \nabla \bar{y}(x) \rangle, \quad \text{a.e. } x \in \Omega, \quad i = 0, 1. \quad (4.78)$$

The proof is omitted here. Also, one could derive (at least formally) the optimality system for the problem similar to the minimization problem.

## 5 A two-dimensional example

In this section, we present a two-dimensional example of Problem  $(\underline{\Lambda}^H[0, 1])$ ; Namely,  $\alpha = 0, \beta = 1$  and according to (1.24), we should assume that neither  $A_0 \leq A_1$  nor  $A_0 \geq A_1$  holds. Since both  $A_0$  and  $A_1$  are positive definite, making a change of variables if necessary, without loss of generality, we may assume that

$$A_0 = \begin{pmatrix} \mu_0 & 0 \\ 0 & \mu_1 \end{pmatrix}, \quad A_1 = I, \quad (5.1)$$

with  $0 < \mu_0 < 1 < \mu_1$ . Recall that this example is comparable with Example 3.3. Let  $\underline{a}(\cdot) \in \overline{\mathcal{A}[0, 1]}^H$  be an optimal control of Problem  $(\underline{\Lambda}^H[0, 1])$ . Then the following holds:

$$\langle \underline{a}(x) \nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \geq \langle B^{-1} \underline{a}(x) \nabla \underline{y}(x), \underline{a}(x) \nabla \underline{y}(x) \rangle, \quad B = A_0, I, \quad \text{a.e. } x \in \Omega. \quad (5.2)$$

To determine an optimal control  $\underline{a}(\cdot)$ , let us make an observation. For given  $x \in \mathbb{R}^n$ , if we denote  $\xi = \nabla \underline{y}(x)$  and  $\bar{A} = \underline{a}(x)$ , then (5.2) reads

$$\langle \bar{A}\xi, \xi \rangle \geq \langle B^{-1}\bar{A}\xi, \bar{A}\xi \rangle, \quad \forall B \in \{A_0, I\}. \quad (5.3)$$

Or, equivalently,

$$\langle \bar{A}\xi, \xi \rangle \geq \langle A_0^{-1}\bar{A}\xi, \bar{A}\xi \rangle, \quad \langle \bar{A}\xi, \xi \rangle \geq \langle \bar{A}\xi, \bar{A}\xi \rangle. \quad (5.4)$$

Hence, it is natural to consider the following problem.

**Problem (P<sub>ξ</sub>)**. For  $\xi \in \mathbb{R}^2$ , find an  $\bar{A} \in \overline{\{A_0, I\}}^H$  such that (5.3) holds.

Here, we recall that

$$\overline{\{A_0, I\}}^H = \left\{ A \in M[\mu_0, \mu_1] \mid \chi_\Omega(\cdot)A \in \overline{L^\infty(\Omega; \{A_0, I\})}^H \right\}.$$

For the above problem, we have the following interesting proposition.

### Proposition 5.1

(i) If  $\bar{A} \in \overline{\{A_0, I\}}^H$  is a solution to Problem (P<sub>ξ</sub>), then

$$\langle \bar{A}\xi, \xi \rangle \geq \langle B^{-1}\bar{A}\xi, \bar{A}\xi \rangle, \quad \forall B \in \overline{\{A_0, I\}}^H. \quad (5.5)$$

(ii) If both  $\bar{A}$  and  $\bar{B}$  are solutions of Problem (P<sub>ξ</sub>), then  $\bar{A}\xi = \bar{B}\xi$ .

**Proof** (i) For any  $B \in \overline{\{A_0, I\}}^H$ , we can find a sequence  $\Omega_k \subseteq \Omega$  such that

$$\chi_{\Omega_k^c}(\cdot)A_0 + \chi_{\Omega_k}(\cdot)I \xrightarrow{H} \chi_\Omega(\cdot)B, \quad \frac{|\Omega_k|}{|\Omega|} \rightarrow \theta,$$

for some  $\theta \in [0, 1]$ . By Property (vi) in Sect. 4.1,

$$B^{-1} \leq (1 - \theta)A_0^{-1} + \theta I.$$

Thus, (5.5) follows.

(ii) Note that

$$\langle \bar{A}\xi, \xi \rangle \geq \langle \bar{B}^{-1}\bar{A}\xi, \bar{A}\xi \rangle, \quad \langle \bar{B}\xi, \xi \rangle \geq \langle \bar{A}^{-1}\bar{B}\xi, \bar{B}\xi \rangle,$$

we have

$$\langle (\bar{A}^{-1} + \bar{B}^{-1})(\bar{A} - \bar{B})\xi, (\bar{A} - \bar{B})\xi \rangle = \langle \bar{B}^{-1}\bar{A}\xi, \bar{A}\xi \rangle - \langle \bar{A}\xi, \xi \rangle + \langle \bar{A}^{-1}\bar{B}\xi, \bar{B}\xi \rangle - \langle \bar{B}\xi, \xi \rangle \leq 0.$$

Therefore, it must hold that  $\bar{A}\xi = \bar{B}\xi$ .  $\square$

The above tells us that to meet the necessary conditions for optimal controls of Problem  $(\underline{A}^H[0, 1])$  on  $\overline{\mathcal{A}[0, 1]}^H$ , it suffices to find, for almost each  $x \in \Omega$ , an  $\bar{A} \in \overline{\{A_0, I\}}^H$  satisfying

$$\langle \bar{A}\nabla \underline{y}(x), \nabla \underline{y}(x) \rangle \geq \langle B^{-1}\bar{A}\nabla \underline{y}(x), \bar{A}\nabla \underline{y}(x) \rangle, \quad B = A_0, I. \quad (5.6)$$

Part (ii) of above proposition means that although  $\bar{A}$  might not be unique,  $\bar{A}\xi$  is unique. Thus, if one can solve Problem (P<sub>ξ</sub>) successfully for each  $\xi \in \mathbb{R}^2$ , then we obtain a map  $\bar{A} : \mathbb{R}^2 \rightarrow \mathbb{S}^2$ . Then

$$\underline{a}(x) = \bar{A}(\nabla \underline{y}(x)), \quad x \in \Omega \quad (5.7)$$

gives an optimal control, where  $\underline{y}(\cdot)$  is a solution to the closed-loop system:

$$\begin{cases} -\nabla \cdot (\bar{A}(\nabla \underline{y}(x)) \nabla \underline{y}(x)) = \underline{\lambda} \underline{y}(x), & x \in \Omega, \\ \underline{y}|_{\partial\Omega} = 0. \end{cases} \quad (5.8)$$

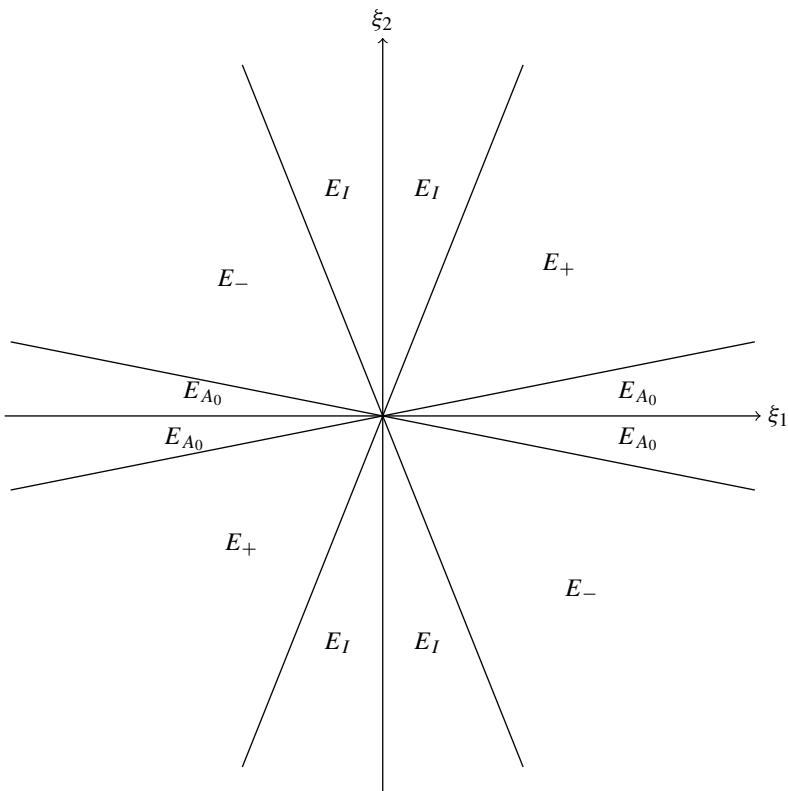
We will see later it is a nonlinear eigenvalue problem.

Now, for given  $\xi = (\xi_1, \xi_2)^T \neq 0$ , we try to find a solution  $\bar{A}$  of Problem  $(\mathbf{P}_\xi)$ .

Let us introduce the following partition of  $\mathbb{R}^2$ :

$$\begin{aligned} E_{A_0} &= \left\{ (\xi_1, \xi_2)^T \neq 0 \mid \xi_2^2 \leq \frac{\mu_0(1 - \mu_0)}{\mu_1(\mu_1 - 1)} \xi_1^2 \right\}, \\ E_I &= \left\{ (\xi_1, \xi_2)^T \neq 0 \mid \xi_2^2 \geq \frac{\mu_1(1 - \mu_0)}{\mu_0(\mu_1 - 1)} \xi_1^2 \right\}, \\ E_+ &= \left\{ (\xi_1, \xi_2)^T \in \mathbb{R}^2 \mid \frac{\mu_0(1 - \mu_0)}{\mu_1(\mu_1 - 1)} \xi_1^2 < \xi_2^2 < \frac{\mu_1(1 - \mu_0)}{\mu_0(\mu_1 - 1)} \xi_1^2, \quad \xi_1 \xi_2 > 0 \right\}, \\ E_- &= \left\{ (\xi_1, \xi_2)^T \in \mathbb{R}^2 \mid \frac{\mu_0(1 - \mu_0)}{\mu_1(\mu_1 - 1)} \xi_1^2 < \xi_2^2 < \frac{\mu_1(1 - \mu_0)}{\mu_0(\mu_1 - 1)} \xi_1^2, \quad \xi_1 \xi_2 < 0 \right\}. \end{aligned} \quad (5.9)$$

These sets are illustrated in the following figure. Clearly,  $E_{A_0}$ ,  $E_I$ ,  $E_+$ ,  $E_-$  are non-empty, mutually disjoint and  $\mathbb{R}^2 = E_{A_0} \cup E_I \cup E_+ \cup E_- \cup \{0\}$ . Note that  $E_{A_0} \cup \{0\}$  and  $E_I \cup \{0\}$  are closed, and  $E_{\pm}$  are open.



With a little calculation, one can see the following:

- $A_0$  is a solution of Problem  $(\mathbf{P}_\xi)$  if and only if  $\langle A_0 \xi, \xi \rangle \geq \langle A_0 \xi, A_0 \xi \rangle$ , i.e.,  $\xi \in E_{A_0}$ .

- $I$  is a solution of Problem  $(\mathbf{P}_\xi)$  if and only if  $\langle \xi, \xi \rangle \geq \langle A_0^{-1}\xi, \xi \rangle$ , i.e.,  $\xi \in E_I$ .
- Neither  $A_0$  nor  $I$  is a solution of Problem  $(\mathbf{P}_\xi)$  if and only if  $\xi \in E_+ \cup E_-$ , i.e.,

$$\langle A_0\xi, \xi \rangle < \langle A_0\xi, A_0\xi \rangle, \quad \langle \xi, \xi \rangle < \langle A_0^{-1}\xi, \xi \rangle. \quad (5.10)$$

Now, let  $\xi = (\xi_1, \xi_2)^T \in E_+ \cup E_-$ , by definition,  $\xi_1, \xi_2 \neq 0$ . Let  $\eta = \bar{A}\xi$ . Since  $\bar{A} \in \overline{\{A_0, I\}}^H \setminus \{A_0, I\}$ , by the proof of Proposition 5.1, (i), there exists a  $\gamma \in (0, 1)$  such that

$$\bar{A}^{-1} \leq (1 - \gamma)A_0^{-1} + \gamma I. \quad (5.11)$$

Then

$$\langle \bar{A}^{-1}\eta, \eta \rangle \leq (1 - \gamma) \langle A_0^{-1}\eta, \eta \rangle + \gamma \langle \eta, \eta \rangle.$$

On the other hand, (5.4) implies

$$\langle \bar{A}^{-1}\eta, \eta \rangle \geq \langle A_0^{-1}\eta, \eta \rangle, \quad \langle \bar{A}^{-1}\eta, \eta \rangle \geq \langle \eta, \eta \rangle.$$

Thus, it should hold that

$$\langle \bar{A}^{-1}\eta, \eta \rangle = \langle A_0^{-1}\eta, \eta \rangle = \langle \eta, \eta \rangle, \quad (5.12)$$

which coincides with (4.72). From the second equality in (5.12), a direct calculation shows

$$\eta = C_\xi \left( \frac{\varepsilon \sqrt{1-s}}{\sqrt{s}} \right), \quad (5.13)$$

with  $C_\xi \neq 0$  and  $\varepsilon = \pm 1$ , where

$$s = \frac{(1 - \mu_0)\mu_1}{\mu_1 - \mu_0} \in (0, 1). \quad (5.14)$$

Let us now determine  $C_\xi$ ,  $\varepsilon$  and  $\gamma$ . We have

$$C_\xi^2 = \langle \eta, \eta \rangle = \langle \bar{A}^{-1}\eta, \eta \rangle = \langle \xi, \eta \rangle = C_\xi \left( \left( \frac{\varepsilon \sqrt{1-s}}{\sqrt{s}} \right), \xi \right).$$

Thus

$$C_\xi = \varepsilon \xi_1 \sqrt{1-s} + \xi_2 \sqrt{s} \quad (5.15)$$

and

$$\begin{aligned} \bar{A}\xi &= \eta = C_\xi \left( \frac{\varepsilon \sqrt{1-s}}{\sqrt{s}} \right) = \left( \frac{\varepsilon \sqrt{1-s}}{\sqrt{s}} \right) \left( \frac{\varepsilon \sqrt{1-s}}{\sqrt{s}} \right)^T \xi \\ &= \left( \frac{1-s}{\varepsilon \sqrt{s(1-s)}} \frac{\varepsilon \sqrt{s(1-s)}}{s} \right) \xi \equiv G\xi. \end{aligned} \quad (5.16)$$

On the other hand, by (5.11) and (5.12), we have

$$\left| \left( (1 - \gamma)A_0^{-1} + \gamma I - \bar{A}^{-1} \right)^{\frac{1}{2}} \eta \right|^2 = \left\langle \left( (1 - \gamma)A_0^{-1} + \gamma I - \bar{A}^{-1} \right) \eta, \eta \right\rangle = 0.$$

This implies

$$\left( (1 - \gamma)A_0^{-1} + \gamma I - \bar{A}^{-1} \right) \eta = 0. \quad (5.17)$$

Thus,

$$\xi = \bar{A}^{-1}\eta = \left( (1 - \gamma)A_0^{-1} + \gamma I \right) \eta = C_\xi \begin{pmatrix} \varepsilon \left( \frac{1-\gamma}{\mu_0} + \gamma \right) \sqrt{1-s} \\ \left( \frac{1-\gamma}{\mu_1} + \gamma \right) \sqrt{s} \end{pmatrix}. \quad (5.18)$$

Hence,  $\operatorname{sgn}(C_\xi) = \operatorname{sgn}(\xi_2)$ , which can be obtained from (5.15) and  $x \in E_+ \cup E_-$  too. Consequently,

$$\varepsilon = \operatorname{sgn}(\xi_1 \xi_2). \quad (5.19)$$

Moreover, it follows from  $\xi_2 = C_\xi \left( \frac{1-\gamma}{\mu_1} + \gamma \right) \sqrt{s}$  that

$$\gamma = \frac{|\xi_2| \sqrt{\frac{\mu_1}{1-\mu_0}} - |\xi_1| \sqrt{\frac{\mu_0}{\mu_1-1}}}{|\xi_1| \sqrt{\mu_0(\mu_1-1)} + |\xi_2| \sqrt{\mu_1(1-\mu_0)}}. \quad (5.20)$$

One can verify that the above  $\gamma$  belonging to  $(0, 1)$  is equivalent to  $\xi \in E_+ \cup E_-$ . Actually, it is a one-to-one mapping from  $\{(\xi_1, \xi_2)^\top \in S^1 \cap (E_+ \cup E_-) | \xi_1 > 0, \xi_2 > 0\}$  to  $(0, 1)$ .

Next, by (ii) of Proposition 5.1,  $\eta \equiv \bar{A}\xi$  only depends on  $\xi$  (independent of the solution  $\bar{A}$ ). We now find an  $\bar{A} \in \overline{\{A_0, I\}}^H$  such that  $\bar{A}\xi = \eta$ . To this end, we try to find an solution in  $\Gamma(A_0, I)$ . By careful calculation, we find such a solution  $\bar{A}$  as the following:

$$\bar{A} = I + (1 - \gamma)Q^{-1}, \quad Q = (A_0 - I)^{-1} + \gamma(I - G). \quad (5.21)$$

Denote  $H = I - G$ . Then  $H \in \mathbb{S}^n$  with  $H \geq 0$  and  $\operatorname{tr}(H) = 1$ . Thus  $\bar{A} \in \Gamma(A_0, I)$ . Moreover,  $H^2 = H$ . Let us verify  $\bar{A}$  defined by (5.21) really satisfies  $\bar{A}\xi = \eta$ . We have

$$(A_0 - I)QH = H + \gamma(A_0 - I)H.$$

Thus

$$\begin{aligned} & A_0^{-1}(A_0 - I)Q(\bar{A}\xi - \eta) \\ &= A_0^{-1}(A_0 - I)Q(H + (1 - \gamma)Q^{-1})\xi \\ &= A_0^{-1}(H + \gamma(A_0 - I)H + (1 - \gamma)(A_0 - I))\xi \\ &= A_0^{-1}(A_0 - (1 - \gamma)G - \gamma A_0 G)\xi = \xi - (1 - \gamma)A_0^{-1}\eta - \gamma\eta = 0. \end{aligned}$$

Thus, along with  $G, \gamma, \varepsilon$  being given by (5.16), (5.19) and (5.20), (when  $\xi \in E_+ \cup E_-$ ) we could get  $\bar{A}$  by (5.21). Generally, we can choose

$$\bar{A}(\xi) = \begin{cases} I, & \xi = 0, \\ A_0, & \xi \in E_{A_0}, \\ I, & \xi \in E_I, \\ I + (1 - \gamma)\left((A_0 - I)^{-1} + \gamma(I - G)\right)^{-1}, & \xi \in E_+ \cup E_-. \end{cases} \quad (5.22)$$

Therefore (5.8) is a nonlinear eigenvalue problem. We now simplify it. Noting (5.16), we see that

$$\underline{a}(x)\nabla \underline{y}(x) = \bar{A}(\nabla \underline{y}(x))\nabla \underline{y}(x) = \mathbf{F}(\nabla \underline{y}(x)), \quad \text{a.e. } x \in \Omega, \quad (5.23)$$

where  $\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is defined as

$$\mathbf{F}(\xi) = \begin{cases} 0, & \xi = 0, \\ A_0\xi, & \xi \in E_{A_0}, \\ \xi, & \xi \in E_I, \\ G_+\xi, & \xi \in E_+, \\ G_-\xi, & \xi \in E_-, \end{cases} \quad (5.24)$$

with

$$G_{\pm} = \begin{pmatrix} 1-s & \pm\sqrt{s(1-s)} \\ \pm\sqrt{s(1-s)} & s \end{pmatrix}. \quad (5.25)$$

Since  $\mathbf{F}(\nabla \underline{y}(x)) = \underline{a}(x) \nabla \underline{y}(x)$ ,  $\underline{y}(\cdot)$  solves

$$\begin{cases} -\nabla \cdot (\mathbf{F}(\nabla \underline{y}(x))) = \underline{\lambda} \underline{y}(x), & \text{in } \Omega, \\ \underline{y}|_{\partial\Omega} = 0. \end{cases} \quad (5.26)$$

Although  $G_{\pm}$  are singular (0 is an eigenvalue of  $G_{\pm}$ ), noting that  $\mathbf{F}(\xi) = \bar{A}(\xi)\xi$ , we still have

$$\mu_0|\xi|^2 \leq \langle \mathbf{F}(\xi), \xi \rangle \leq \mu_1|\xi|^2, \quad \forall \xi \in \mathbb{R}^n. \quad (5.27)$$

On the other hand, it is not difficult to verify that  $\langle \mathbf{F}(\cdot), \cdot \rangle$  is convex in  $\mathbb{R}^n$  and

$$|\mathbf{F}(\xi) - \mathbf{F}(\tilde{\xi})| \leq \mu_1|\xi - \tilde{\xi}|, \quad \forall \xi, \tilde{\xi} \in \mathbb{R}^n. \quad (5.28)$$

Consequently, there is  $\tilde{y}(\cdot) \in W_0^{1,2}(\Omega)$  such that

$$\tilde{\lambda} \equiv \frac{\int_{\Omega} \langle \mathbf{F}(\nabla \tilde{y}(x)), \nabla \tilde{y}(x) \rangle dx}{\int_{\Omega} \|\tilde{y}(x)\|^2 dx} = \inf_{\substack{y(\cdot) \in W_0^{1,2}(\Omega) \\ y(\cdot) \neq 0}} \frac{\int_{\Omega} \langle \mathbf{F}(\nabla y(x)), \nabla y(x) \rangle dx}{\int_{\Omega} \|y(x)\|^2 dx}.$$

We have

$$\begin{cases} -\nabla \cdot (\mathbf{F}(\nabla \tilde{y}(x))) = \tilde{\lambda} \tilde{y}(x), & \text{in } \Omega, \\ \tilde{y}|_{\partial\Omega} = 0. \end{cases}$$

Moreover, let  $\tilde{a}(\cdot) = \bar{A}(\tilde{y}(\cdot))$ . Then  $\tilde{a}(\cdot) \in \overline{\mathcal{A}[0, 1]}^H$  and

$$\begin{cases} -\nabla \cdot (\tilde{a}(x) \nabla \tilde{y}(x)) = \tilde{\lambda} \tilde{y}(x), & \text{in } \Omega, \\ \tilde{y}|_{\partial\Omega} = 0, \end{cases}$$

which implies  $\underline{\lambda} \leq \lambda_{\tilde{a}(\cdot)} \leq \tilde{\lambda} \leq \underline{\lambda}$ . This implies

$$\underline{\lambda} = \frac{\int_{\Omega} \langle \mathbf{F}(\nabla \underline{y}(x)), \nabla \underline{y}(x) \rangle dx}{\int_{\Omega} \|\underline{y}(x)\|^2 dx} = \inf_{\substack{y(\cdot) \in W_0^{1,2}(\Omega) \\ y(\cdot) \neq 0}} \frac{\int_{\Omega} \langle \mathbf{F}(\nabla y(x)), \nabla y(x) \rangle dx}{\int_{\Omega} \|y(x)\|^2 dx}. \quad (5.29)$$

Therefore,  $\underline{y}(\cdot)$  is an optimal state of Problem  $(\underline{\Lambda}^H[0, 1])$  if and only if it is a minimizer of

$$\mathcal{F}(y(\cdot)) = \frac{\int_{\Omega} \langle \mathbf{F}(\nabla y(x)), \nabla y(x) \rangle dx}{\int_{\Omega} \|y(x)\|^2 dx}$$

over  $W_0^{1,2}(\Omega) \setminus \{0\}$ .

The results of this section can be summarized as follow: To get a solution of Problem  $(\underline{\Lambda}^H[0, 1])$ , one can first find a nontrivial solution of (5.26) with the smallest positive

number  $\underline{\lambda}$ , or equivalently, find a minimizer of  $\mathcal{F}(\cdot)$  over  $W_0^{1,2}(\Omega) \setminus \{0\}$ . Then define  $\underline{a}(x) = \bar{A}(\nabla \underline{y}(x))$ , getting a solution of Problem  $(\underline{\Delta}^H[0, 1])$ .

## 6 Concluding remarks

We have investigated the maximization and minimization problems of the principal eigenvalue of elliptic operators with the Dirichlet boundary condition. The control appears in the diffusion matrix (the leading coefficient). These problems are well-motivated by composite material design to optimize the heat conduct property of the material (cooling down as quick as possible, or preserving the temperature as long as possible). For maximization problem, due to the concavity of the principle eigenvalue as a functional of the leading coefficient, as long as the control set is convex, optimal control and its characterization can be obtained easily. When the control set is not convex, we introduce the usual convexification to guarantee the existence of an optimal relaxed control. Then some necessary conditions can also be obtained. From an example, we see that uniformly mixing two materials might not be optimal in the maximization problem.

For minimization problem, the situation is much more complicated due to the concavity of the principle eigenvalue as a functional of the control. We adopt the  $H$ -convergence so that the existence of the  $H$ -relaxed optimal control could be guaranteed. Instead of looking at the most general situation, we concentrate on the case of the lamination of two materials whose diffusibility matrices are given. Some interesting necessary conditions are derived. It is worthy of pointing out that even both two materials have their diagonal diffusibility matrices, the optimal diffusibility matrix could be non-diagonal. Such a situation has been exhibited through an illustrative example in Section 5.

**Acknowledgements** The authors would like to thank the referee for his/her comments and suggestions.

## References

1. Adams, R.A.: Sobolev Spaces. Academic Press, New York (1975)
2. Allaire, G.: Homogenization and two-scale convergence. SIAM J. Math. Anal. **23**, 1482–1518 (1992)
3. Allaire, G.: Shape Optimization by the Homogenization Method. Springer, New York (2002)
4. Ansini, N., Dal Maso, G., Zeppieri, C.I.:  $\Gamma$ -convergence and  $H$ -convergence of linear elliptic operators. J. Math. Pures Appl. **99**, 321–329 (2013)
5. Belhachmi, Z., Bucur, D., Buttazzo, G., Sac-Epee, J.-M.: Shape optimization problems for eigenvalues of elliptic operators. Z. Angew. Math. Mech. **86**, 171–184 (2006)
6. Bucur, D., Buttazzo, G., Nitsch, C.: Two optimization problems in thermal insulation. Not. AMS **64**, 830–835 (2017)
7. Buttazzo, G., Dal Maso, G.: Shape optimization for Dirichlet problems: relax formulation and optimality conditions. Appl. Math. Optim. **23**, 17–49 (1991)
8. Casado-Díaz, J.: Smoothness properties for the optimal mixture of two isotropic materials: the compliance and eigenvalue problems. SIAM J. Control Optim. **53**, 2319–2349 (2015)
9. Casado-Díaz, J., Couce-Calvo, J., Martín-Gómez, J.D.: Relaxation of a control problem in the coefficients with a functional of quadratic growth in the gradient. SIAM J. Control Optim. **47**, 1428–1459 (2008)
10. Caffarelli, L.A., Lin, F.H.: An optimal partition problem for eigenvalues. J. Sci. Comput. **31**, 5–18 (2007)
11. Cherkaev, A.: Variational Methods for Structural Optimization. Springer, New York (2000)
12. Cox, S.J., Kawohl, B., Uhlig, P.X.: On the optimal insulation of conductors. J. Optim. Theory Appl. **100**, 253–263 (1999)
13. Cox, S.J., McLaughlin, J.R.: Extremal eigenvalue problems for composite membranes, I. II. Appl. Math. Optim. **22**(153–167), 169–187 (1990)

14. Cuccu, F., Emamizadeh, B., Porru, G.: Optimization of the first eigenvalue in problems involving the  $p$ -Laplacian. *Proc. AMS* **137**, 1677–1687 (2009)
15. Gilbarg, D., Trudinger, N.S.: *Elliptic Partial Differential Equations of Second Order*, 2nd edn. Springer, Berlin (1998). (Theorem 8.38)
16. Gallouet, T., Monier, A.: On the regularity of solutions to elliptic equations. *Rend. Mat. Appl.* **19**, 471–488 (2000)
17. Gustafsson, B., Mossino, J.: A note on  $H$ -convergenc, [arXiv: math/0608286v1](https://arxiv.org/abs/math/0608286v1), 11 (2006)
18. Henrot, A.: *Extremum Problems for Eigenvalues of Elliptic Operators*. Birkhäuser, Boston (2006)
19. Húska, J., Poláčik, P., Safonov, M.V.: Principal eigenvalues, spectral gaps and exponential separation between positive and sign-changing solutions of parabolic equations. *Dis. Contin. Dyn. Syst.* **2005**(suppl), 427–435 (2005)
20. Li, B., Lou, H.: Cesari-type conditions for semilinear elliptic equations with leading term containing controls. *Math. Control Relat. Fields* **1**, 41–59 (2011)
21. Li, B., Lou, H., Xu, Y.: Relaxation of optimal control problem governed by semilinear elliptic equation with leading term containing controls. *Acta Appl. Math.* **130**, 205–236 (2014)
22. Li, X., Yong, J.: *Optimal Control Theory for Infinite Dimensional Systems*. Birkhäuser, Boston (1995)
23. Lou, H.: Optimality conditions for semilinear parabolic equations with controls in leading term. *ESAIM Control Optim. Calc. Var.* **17**, 975–994 (2011)
24. Lou, H., Yong, J.: Optimality conditions for semilinear elliptic equations with leading term containing controls. *SIAM J. Control Optim.* **48**, 2366–2387 (2009)
25. Lou, H., Yong, J.: Second-order necessary conditions for optimal control of semilinear elliptic equations with leading term containing controls. *Math. Control Relat. Fields* **8**, 57–88 (2018)
26. Meyers, N.G.: An  $L^p$ -estimate for the gradient of solutions of second order elliptic divergence equations. *Ann. Scuola Norm. Sup. Pisa* **17**, 189–206 (1963)
27. Münch, A., Pedregal, P., Periago, F.: Relaxation of an optimal design problem for the heat equation. *J. Math. Pures Appl.* **89**, 225–247 (2008)
28. Murat, F., Tartar, L.: On the control of coefficients in partial differential equations. In: Cherkaev, A., Kohn, R. (eds.) *Topics in the Mathematical Modelling of Composite Materials*, pp. 1–8. Birkhäuser, Boston (1997). (Originally published in *Lecture Notes in Economics and Mathematical Systems Modelling*, Springer-Verlag, 420–426 (1975))
29. Murat, F., Tartar, L.:  $H$ -convergence. In: Cherkaev, A., Kohn, R. (eds.) *Topics in the Mathematical Modelling of Composite Materials*, pp. 21–43. Birkhäuser, Boston (1997). (Originally published in *Séminar d'Analyse Fonctionnelle et Numérique de l'Université d'Alger*, (1977))
30. Murat, F., Tartar, L.: Calculus of variations and homogenization. In: Cherkaev, A., Kohn, R. (eds.) *Topics in the Mathematical Modelling of Composite Materials*, pp. 139–173. Birkhäuser, Boston (1997). (Originally published in *Eyrrolles*, 319–369 (1985))
31. Spagnolo, S.: Sulla convergenza di soluzioni di equazioni paraboliche ed ellittiche. *Ann. Sc. Norm. Sup. Pisa* **22**, 571–597 (1968) (In Italian)
32. Tartar, L.: Quelques remarques sur l'homogénéisation. In: *Proceedings of the Japan–France Seminar. "Functional Analysis and Numerical Analysis"*, Japan Society for the Promotion of Sciences, pp. 469–482 (1976)
33. Tartar, L.: Estimations of homogenized coefficients. In: Cherkaev, A., Kohn, R. (eds.) *Topics in the Mathematical Modelling of Composite Materials*, pp. 9–20. Birkhäuser, Boston (1997). (Originally published in *Computing Methods in Applied Sciences and Engineering, Lecture Notes in Math. 704*, Springer-Verlag, 364–373 (1977))
34. Vrdoljak, M.: Classical optimal design in two-phase conductivity problems. *SIAM J. Control Optim.* **54**, 2020–2035 (2016)