Numerical approximation of a one-dimensional elliptic optimal design problem

J. Casado-Díaz[1] C. Castro[2] M. Luna-Laynez[1] E. Zuazua[3,4]

December 22, 2010

Abstract

We address the numerical approximation by finite element methods of an optimal design problem for a two phase material in one space dimension. This problem, in the continuous setting, due to high frequency oscillations, often has not a classical solution and a relaxed formulation is needed to ensure existence. By the contrary, the discrete versions obtained by numerical approximation have a solution. In this article we prove the convergence of the discretizations and obtain convergence rates. We also show a faster convergence when the relaxed version of the continuous problem is taken into account when building the discretization strategy. In particular it is worth emphasazing that, even when the original problem has a classical solution so that relaxation is not necessary, numerical algorithms converge faster when implemented on the relaxed version.

Key words. Control in the coefficients, composite optimal design, relaxation, numerical approximation, finite elements.

AMS subject classification. 49M25, 49J20.

1 Introduction

This paper is devoted to the finite element numerical analysis of a problem of optimal mixture of two (thermal or electrical) materials in order to minimize a given functional in one space dimension.

Let Ω be a bounded open set of $\mathbb{R}^N$, $N \geq 1$ (although our analysis is limited to the case $N = 1$, the problem makes sense in any space dimension) and consider the following optimization problem,

$$\begin{cases} \text{Find } \omega_0 \in \mathcal{U} \text{ such that} \\ \mathcal{J}(\omega_0) = \min_{\omega \in \mathcal{U}} \mathcal{J}(\omega). \end{cases} \tag{1.1}$$

Here ω, the control, is a measurable subset of Ω, $\mathcal{J}(\omega)$, the cost functional, is of the form,

$$\mathcal{J}(\omega) = \int_{\omega} F_1(x, u, \nabla u)\, dx + \int_{\Omega \setminus \omega} F_2(x, u, \nabla u)\, dx, \tag{1.2}$$

[1]Dpto. de Ecuaciones Diferenciales y Análisis Numérico, Facultad de Matemáticas, Universidad de Sevilla, C. Tarfía s/n, 41012 Sevilla, Spain. jcasadod@us,es, mllaynez@us.es

[2]Dpto. de Matemáticas e Informática, ETSI caminos, canales y puertos, Universidad Politécnica de Madrid, Ciudad Universitaria, 28040 Madrid, Spain. carlos.castro@upm.es

[3]Basque Center for Applied Mathematics (BCAM), Bizkaia Technology Park, Building 500. E-48160 Derio - Basque Country - Spain. zuazua@bcamath.org

[4]IKERBASQUE, Basque Foundation for Science, E-48011 Bilbao - Basque Country - Spain.

where $F_1, F_2 : \Omega \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ are given functions, and u, the state, is the solution of,

$$\begin{cases} -\mathrm{div}\Big((\alpha\chi_\omega + \beta(1-\chi_\omega))\,\nabla u\Big) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{1.3}$$

for some given source term $f : \Omega \to \mathbb{R}$. The positive constants α, β represent the two materials, determining the coefficients of the corresponding diffusion matrices. Some restrictions can and have to be imposed to the control ω depending on the problem. For example, an interesting case is when the material α is more efficient than the material β but it is also more expensive. Then, it is usual to consider a restriction of the form $|\omega| \leq \kappa$, limiting the use of the material α. We include this restriction in the admissible set of controls $\mathcal{U}$,

$$\mathcal{U} = \{\omega \subset \Omega \,:\, \omega \text{ measurable }, \quad |\omega| \leq \kappa\}\,. \tag{1.4}$$

The existence of an optimal set ω fulfilling these constraints, for which the function u solution of (1.3) minimizes $\mathcal{J}$ does not hold in general ([13], [14]). In these cases, it is natural to look for minimizing sequences, i.e. sequences $\{\omega_l\}_{l=1}^{\infty} \subset \mathcal{U}$ such that

$$\lim_{l\to\infty} \mathcal{J}(\omega_l) = \inf_{\omega\in\mathcal{U}} \mathcal{J}(\omega),$$

since they provide near optimal designs. A usual procedure to find such sequences is to introduce a relaxed version of the problem for which a minimizer exists. Then, a suitable approximation of the minimizers provides minimizing sequences of the original problem.

For a sequentially continuous functional $\mathcal{J}$, in the weak topology of the Sobolev space $H^1(\Omega)$ ([1], [11], [18]), this relaxation can be obtained by replacing in equation (1.3) the function χ_ω by a measurable function θ taking its values in the closed interval $[0, 1]$ and the function $(\alpha\chi_\omega + \beta(1-\chi_\omega))$ by a matrix function A in the set $\mathcal{K}(\theta)$ of matrices constructed by homogenization (see e.g. [15], [17], [19]) mixing the materials α and β with respective proportions θ and $1-\theta$. Remark that the set $\mathcal{K}(\theta)$ is known in the case described above, corresponding to the mixture of two isotropic materials ([12], [20]), but not in other interesting cases such as the mixture of more than two materials, anisotropic materials... Henceforth we denote by $\hat{\mathcal{U}}$ the set of relaxed controls (θ, A).

Note that functionals of the form (1.2) are not sequentially continuous in the weak topology of $H^1(\Omega)$, in general. In those cases, to obtain the relaxed version ([4]) we must replace the set of controls χ_ω and coefficients $(\alpha\chi_\omega + \beta(1-\chi_\omega))$ by the pairs $(\theta, A) \in \hat{\mathcal{U}}$ as above, and the functional $\mathcal{J}$ by another one of the form

$$\hat{\mathcal{J}}(\theta, A) = \int_\Omega H(x, u, \nabla u, A\nabla u, \theta)\, dx. \tag{1.5}$$

where u is solution of the homogenized problem,

$$\begin{cases} -\mathrm{div}\Big(A\nabla u\Big) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases} \tag{1.6}$$

An explicit expression of the function H is only known in some particular cases ([2], [4], [5], [6], [9], [10], [16], [21]). It satisfies $H(x, u, \nabla u, A\nabla u, \theta) = F_1(x, u, \nabla u)\chi_\omega + F_2(x, u, \nabla u)\chi_{\Omega\setminus\omega}$ if $\theta = \chi_\omega$ and $A = (\alpha\chi_\omega + \beta(1-\chi_\omega))\, I$ and, so, the relaxed functional is in fact an extension of the original one to the larger set of relaxed controls. The relaxed control problem reads

$$\begin{cases} \text{Find } (\theta_0, A_0) \in \hat{\mathcal{U}} \text{ such that} \\ \hat{\mathcal{J}}(\theta_0, A_0) = \min_{(\theta,A)\in\hat{\mathcal{U}}} \hat{\mathcal{J}}(\theta, A). \end{cases} \tag{1.7}$$

In practical applications, in order to solve numerically the above control problem (1.1), it is necessary to introduce a discretization of both the control set and the functional. In the present context we have at least two approaches to this numerical approximation issue. The one based on the discretization of the original problem and the one relying on the discretization of the relaxed version. Recently, in [6] and [7] both discretization procedures have been shown to converge (in these articles some partially relaxed versions have also been studied in which the class of controls under consideration is enlarged but not to the extent of exhausting the class of the relaxed version of the problem; we refer to [10] for a related result).

In this paper we compare and get convergence rates for the sequences of discrete minimizers obtained with both approximation methods. These issues are addressed in the simplest one-dimensional setting, where the partial differential equation (1.3) is reduced to an ordinary differential equation, the set $\mathcal{K}(\theta)$ is well known to be reduced to the harmonic mean of α and β with respective proportions θ and $1-\theta$ and the function H is explicitly known. Note that in this case we can write $\hat{\mathcal{J}}(\theta, A) = \hat{\mathcal{J}}(\theta)$ in (1.5), since A is completely determined by θ, and $\hat{\mathcal{U}}$ is just the set of measurable functions $\theta : \Omega \to [0,1]$, with integral less or equal than κ.

To make precise our results we first consider the discretization of the set of controls but not of the the state equation (1.3). In the context of finite element approximation methods, we can consider a decomposition of Ω in elements with maximum size r and subsets ω constituted by unions of a subset of such elements. If we denote by $\mathcal{U}^r$ the set of such subsets, the discrete problem reads

$$\begin{cases} \text{Find } \omega_0^r \in \mathcal{U}^r, \text{ such that} \\ \mathcal{J}(\omega_0^r) = \min\limits_{\omega \in \mathcal{U}^r} \mathcal{J}(\omega). \end{cases} \tag{1.8}$$

The discrete space of controls obtained in this way $\mathcal{U}^r$ is compact in the strong topology of $L^1(\Omega)$ and the corresponding state functions are compact in $H^1(\Omega)$. Therefore, the discretized problem has a solution without the need for a relaxed version.

In this way we obtain a sequence of discrete minimizers $\{\omega_0^r\}_r$ that are likely to constitute a minimizing sequence of $\mathcal{J}$ in $\mathcal{U}$, as $r \to 0$. We show that this is the case and we give convergence rates for

$$\mathcal{J}(\omega_0^r) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega), \qquad \text{as } r \to 0. \tag{1.9}$$

On the other hand, instead of discretizing the original control problem we can discretize the relaxed version. After introducing a decomposition of Ω in elements, with maximal size r, we can consider the set $\hat{\mathcal{U}}^r$ of functions $\theta \in \hat{\mathcal{U}}$ which are constant on each element. The discrete relaxed problem reads,

$$\begin{cases} \text{Find } \hat{\theta}_0^r \in \hat{\mathcal{U}}^r \text{ such that} \\ \hat{\mathcal{J}}(\hat{\theta}_0^r) = \min\limits_{\theta \in \hat{\mathcal{U}}^r} \hat{\mathcal{J}}(\theta), \end{cases} \tag{1.10}$$

As above, we show that

$$\hat{\mathcal{J}}(\hat{\theta}_0^r) - \inf_{\omega \in \hat{\mathcal{U}}} \mathcal{J}(u) \to 0, \qquad \text{as } r \to 0 \tag{1.11}$$

and we give convergence rates.

Once a discrete relaxed minimizer is known $\hat{\theta}_0^r$ we can construct a sequence $\{\omega^{k,r}\}_{k=1}^{\infty} \subset \mathcal{U}$ such that

$$\lim_{k \to \infty} \mathcal{J}(\omega^{k,r}) = \hat{\mathcal{J}}(\hat{\theta}_0^r).$$

This provides a minimizing sequence of the original problem. As we show, the sequence $\{\omega^{k,r}\}_{k=1}^{\infty}$ can be constructed explicitly from $\hat{\theta}_0^r$, without almost no computational cost.

Our results show that it is better to discretize the relaxed problem, in the sense that we get a faster convergence rate, as $r \to 0$, for (1.11) than the one obtained for (1.9). This is true even in the case where the original problem has a solution and so, the relaxation is unnecessary from a theoretical point of view. Despite of this, the relaxed version of the original minimization problem can always be formulated and our results show that it is indeed better to approximate numerically the optimal design problem in these cases too.

From a computational point of view, besides of discretizing the set of controls we must also discretize the state equation ((1.3) or (1.6)). This requires a second decomposition of Ω constituted by elements of maximum size h. A natural assumption is to consider this new decomposition as a refinement of the one used for the control set, or vice versa.

In the context of the original unrelaxed control problem, denoting by u^h the P_1-finite element approximation of the solution of (1.3) and defining $\mathcal{J}^h$ by

$$\mathcal{J}^h(\omega) = \int_{\omega} F_1(x, u^h, \nabla u^h)\, dx + \int_{\Omega\setminus\omega} F_2(x, u^h, \nabla u^h)\, dx,$$

the full discrete control problem reads,

$$\begin{cases} \text{Find } \omega_0^{r,h} \in \mathcal{U}^r \text{ such that} \\ \mathcal{J}^h(\omega_0^{r,h}) = \min\limits_{\omega\in\mathcal{U}^r} \mathcal{J}^h(\omega). \end{cases} \tag{1.12}$$

Analogously, we can define a full discretization of the relaxed problem by considering

$$\hat{\mathcal{J}}^h(\theta) = \int_{\Omega} H(x, u^h, \nabla u^h, A\nabla u^h, \theta)\, dx. \tag{1.13}$$

where u^h is the P_1-finite element approximation of (1.6). The fully discrete relaxed problem in this case is

$$\begin{cases} \text{Find } \hat{\theta}_0^{r,h} \in \hat{\mathcal{U}}^r \text{ such that} \\ \hat{\mathcal{J}}^h(\hat{\theta}_0^{r,h}) = \min\limits_{\theta\in\hat{\mathcal{U}}^r} \hat{\mathcal{J}}^h(\theta). \end{cases} \tag{1.14}$$

We focus on the convergence rates for the sequences $\{\omega_0^{r,h}\}_{r,h}$ and $\{\hat{\theta}_0^{r,h}\}_{r,h}$ obtained with the two approaches above respectively. More precisely we compare the sequences

$$\mathcal{J}(\omega_0^{r,h}) - \inf_{\omega\in\mathcal{U}} \mathcal{J}(\omega), \qquad \text{and} \qquad \hat{\mathcal{J}}(\hat{\theta}_0^{r,h}) - \inf_{\omega\in\mathcal{U}} \mathcal{J}(\omega),$$

as $r, h \to 0$.

The following results are proved:

- Discretizing the relaxed formulation we show that, solving the state equation by the P_1-finite element method in a mesh of size h and taking the control θ to be piecewise constant on elements of a coarser mesh of size $\sqrt{h}$, the error is of order h.

  This constitutes a bigrid or multi-scale strategy, implemented on the relaxed version, in the sense that the discretization of the PDE and that of the control are performed on two different grids. The PDE is discretized in the fine grid of size h while the control is discretized in the coarse one of size $\sqrt{h}$.

- Discretizing the original unrelaxed problem, solving the state equation by a P_1-finite element method in a mesh of size h and taking the control χ_ω piecewise constant in the elements of such mesh, we show that the error is of order $h^{1-\varepsilon}$, with ε arbitrarily small if the functions F_i in (1.2) do not depend on the variable u and $\varepsilon = 1/2$ otherwise.

A bigrid strategy but discretizing the PDE in the coarser grid (instead of the finer one) can produce lack of convergence both for the unrelaxed and relaxed problems. In particular, the minimizers for the discrete problem will possibly give a non-minimizing sequence of the continuous control problem, as $r, h \to 0$.

We also give an explicit example in which the functional is independent of u, showing our estimates are nearly sharp. To be more precise, our example shows the optimality of the estimates in the case in which the relaxed version of the problem is discretized, while an order h of convergence is obtained when the original problem is discretized, thus showing that our estimates are nearly optimal.

Therefore the approach based on the discretization of the relaxed formulation provides a better approximation and a faster convergence rate with a lower computational cost. The computational cost and the complexity of this approach is lower since the controls are discretized in a mesh or order $\sqrt{h}$ instead of h. Furthermore, the minimizers for the corresponding discrete optimization problems are easier to find numerically. Indeed, thanks to the convexity of the relaxed control set, gradient like algorithms can be implemented. This is in contrast with the unrelaxed problem where the control set is not convex and we cannot compute variations. Instead, much less efficient methods as Montecarlo or genetic algorithms should be used.

By the contrary, the advantages of discretizing directly the original problem are that, on one hand, one does not need to know the relaxed formulation and, second, it provides a physical control (i.e. a characteristic function) instead of a relaxed one. However, this later drawback can be overcame when dealing with the discretization of the relaxed problem since can approximate the relaxed optimal control by physical ones, with almost not computational cost.

This paper provides a complete analysis of the rate of convergence of the finite element approximation of the optimal design problem under consideration. Whether this classical engineering practice leads to convergent algorithms is unknown in many other optimal design problems, except in some other particular examples as it occurs when dealing with the optimal shape design of the domain for Dirichlet Laplacian in two space dimensions (see [8]). Note however that, in the later, there is no result about the convergence rate.

Although the present article is devoted to the study of the 1-d optimal design problem, some remarks about the N-dimensional case are given in the last section of the paper.

Some definitions and notations:

• For a number $r \in \mathbb{R}$ we denote by $[r]$ the integer part of r.

• For a (Lebesgue) measurable subset E of $(0,1)$, with positive measure, and a function w in $L^1(0,1)$, we denote the mean value of w in E by

$$\fint_E w\,dx = \frac{1}{|E|}\int_E w\,dx.$$

• The set of functions of bounded variation in $(0,1)$ is denoted by $BV(0,1)$. If ψ is in $BV(0,1)$ and I is a subinterval of $[0,1]$, then $V_I(\psi)$ represents the total variation of ψ in I.

• Along the paper, α and β are two positive constants.

• For $p \in [0,1]$, we denote by $M(p) \in \mathbb{R}$ the harmonic mean of α and β with proportions p and $1-p$ respectively, given by

$$M(p) = \left(\frac{p}{\alpha} + \frac{1-p}{\beta}\right)^{-1} = \frac{\alpha\beta}{(1-p)\alpha + p\beta}\ .$$

Note that $M(1) = \alpha$, $M(0) = \beta$, and

$$\alpha \le M(p) \le \beta, \quad \forall p \in [0,1]. \tag{1.15}$$

For every $\theta \in L^\infty(0,1;[0,1])$ we define $M_\theta \in L^\infty(\Omega)$ by

$$M_\theta(x) = M(\theta(x)), \quad \text{for a.e. } x \in (0,1).$$

- For a matrix $A \in \mathbb{R}^{N\times N}$, we denote by $Eig(A)$ the set of its eigenvalues.
- Let Φ be a function defined in the interval $(0,\delta)$, for some $\delta > 0$. The equality $\Phi = o(h)$ (Landau symbol) means

$$\lim_{h\to 0} \frac{\Phi(h)}{h} = 0.$$

- We denote by C a generic positive constant which can change from line to line.

2 Discretization and error estimates

2.1 The main results

In this section we state the main results of the paper. They are referred to the numerical analysis of a control problem for the $1-d$ elliptic state equation in $\Omega = (0,1)$ below, the control being the space-dependent coefficient:

$$\begin{cases} -\dfrac{d}{dx}\left((\alpha\chi_\omega + \beta(1-\chi_\omega))\dfrac{du}{dx}\right) = f \quad \text{in } (0,1) \\ u(0) = u(1) = 0, \end{cases} \tag{2.1}$$

where α and β are two fixed positive constants and f a given function in (at least) $L^1(0,1)$.

Defining, for a fixed constant $\kappa > 0$, the set of admissible controls as (1.4), our aim is to choose $\omega \in \mathcal{U}$ such that the unique solution $u_\omega \in H_0^1(0,1)$ of problem (2.1) minimizes the functional $\mathcal{J} : \mathcal{U} \longrightarrow \mathbb{R}$ defined as the 1-d version of (1.2), i.e.

$$\mathcal{J}(\omega) = \int_\omega F_1\left(x, u_\omega, \frac{du_\omega}{dx}\right)dx + \int_{(0,1)\setminus\omega} F_2\left(x, u_\omega, \frac{du_\omega}{dx}\right)dx, \quad \forall \omega \in \mathcal{U}. \tag{2.2}$$

Here $F_1, F_2 : (0,1)\times\mathbb{R}\times\mathbb{R} \to \mathbb{R}$ satisfy

$$F_i \in W^{1,\infty}((0,1)\times(-R,R)\times(-R,R)), \quad \forall i \in \{1,2\},\ \forall R > 0. \tag{2.3}$$

As we said in the introduction α and β represent two materials which we want to mix in order to minimize $\mathcal{J}$. The constant κ is the maximum quantity of material α that can be used in the mixture. Remark that taking $\kappa \geq 1$ would be equivalent to not imposing any restriction in the set of admissible sets ω.

Remark 2.1 *In (2.1), we consider homogeneous Dirichlet conditions to fix ideas, but our results also hold for non-homogeneous Dirichlet conditions or other boundary conditions such as Fourier or Neumann ones.*

We can also consider the functions F_i satisfying weaker assumptions than (2.3) but then the error estimates we find for the numerical approximations defined below are worse.

It is well known that the original minimization problem (1.1) has not a solution in general ([13], [14]). Therefore, it is necessary to introduce a relaxation. However, as we have mentioned in the introduction, for numerical purposes it is often convenient to work in the relaxed version of the problem even when the original formulation has a minimizer. The relaxed version thus plays a key role in the numerical analysis we develop in this article.

The following result provides a characterization of the relaxation:

Theorem 2.2 *A relaxation of problem (1.1) is given by*

$$\begin{cases} \text{Find } \theta_0 \in \hat{\mathcal{U}} \text{ such that} \\ \hat{\mathcal{J}}(\theta_0) = \min_{\theta \in \hat{\mathcal{U}}} \hat{\mathcal{J}}(\theta), \end{cases} \tag{2.4}$$

where

$$\hat{\mathcal{U}} = \left\{ \theta \in L^\infty(0,1;[0,1]) \, : \, \int_0^1 \theta dx \leq \kappa \right\}, \tag{2.5}$$

and $\hat{\mathcal{J}} : \hat{\mathcal{U}} \longrightarrow \mathbb{R}$ *is defined by*

$$\hat{\mathcal{J}}(\theta) = \int_0^1 \left(\theta F_1 \left(x, u_\theta, \frac{M_\theta}{\alpha} \frac{du_\theta}{dx} \right) + (1-\theta) F_2 \left(x, u_\theta, \frac{M_\theta}{\beta} \frac{du_\theta}{dx} \right) \right) dx, \tag{2.6}$$

for every $\theta \in \hat{\mathcal{U}}$*, with* $u = u_\theta$ *the solution of*

$$\begin{cases} -\dfrac{d}{dx} \left(M_\theta \dfrac{du}{dx} \right) = f \quad \text{in } (0,1) \\ u(0) = u(1) = 0. \end{cases} \tag{2.7}$$

Remark 2.3 *Theorem 2.2 also holds true for every* $f \in H^{-1}(0,1)$ *and more general nonlinearities* F_1, F_2*. Indeed, it is enough to assume that* F_1, F_2 *are two Caratheodory functions (measurable with respect to* x *and continuous with respect to* (s,ξ)*) such that for every* $R > 0$*, the functions* $\varphi_{1,R}$*,* $\varphi_{2,R}$ *defined as*

$$\varphi_{i,R}(x) = \sup_{|s|+|\xi| \leq R} |F_i(x,s,\xi)|, \quad \text{for a.e. } x \in (0,1), \quad \forall i \in \{1,2\},$$

belong to $L^1(0,1)$*.*

Remark 2.4 *For every* $\omega \subset (0,1)$ *measurable, we have*

$$\mathcal{J}(\omega) = \hat{\mathcal{J}}(\chi_\omega).$$

Therefore, $\hat{\mathcal{J}}$ *is in fact an extension of the functional* $\chi_\omega \mapsto \mathcal{J}(\omega)$ *defined on* $L^\infty(0,1;\{0,1\})$ *to the relaxed control set* $L^\infty(0,1;[0,1])$*.*

Remark 2.5 *Theorem 2.2 is a generalization of Proposition 4.1 and Theorem 4.3 in [4], where the multi-dimensional case is also considered.*

In the present paper, we are interested mainly in the numerical analysis of problem (1.1). For this purpose, thanks to Theorem 2.2, two choices are possible: to discretize directly problem (1.1) or two discretize the relaxed problem (2.4). Our goal is to compare these two possibilities.

To this aim, given $r > 0$, we take a partition $\mathcal{P}^r = \{y_k\}_{k=0}^{m_r}$ of $[0,1]$, with $m_r \in \mathbb{N}$, such that

$$r = \max_{1 \leq k \leq m_r} (y_k - y_{k-1}). \tag{2.8}$$

Then, we define $\hat{\mathcal{U}}^r$ and $\mathcal{U}^r$ as the subsets of $\hat{\mathcal{U}}$ given by

$$\hat{\mathcal{U}}^r = \{ \theta \in \hat{\mathcal{U}} \, : \, \theta = \sum_{k=1}^{m_r} t_k \chi_{(y_{k-1}, y_k)} \text{ a.e. in } (0,1), \text{ with } t_k \in [0,1], \ 1 \leq k \leq m_r \} \tag{2.9}$$

$$\mathcal{U}^r = \{ \omega \subset (0,1) \, : \, \chi_\omega \in \hat{\mathcal{U}}^r \}. \tag{2.10}$$

Associated to these subsets we can consider the two discretizations of the control problem given by (1.10) and (1.8).

Note that problem (1.8) is a discretization of the original minimization problem (1.1), while (1.10) is a discretization of the relaxed problem (2.4).

The following theorems provide estimates on the difference between these problems and (2.4). Some versions of Theorem 2.6 can also be obtained in the N-dimensional case, see Section 8.

Theorem 2.6 *Assuming $f \in L^1(0,1)$, problem (1.10) has a solution for every $r > 0$, and we have*

$$0 \leq \min_{\theta \in \hat{\mathcal{U}}^r} \hat{\mathcal{J}}(\theta) - \min_{\theta \in \hat{\mathcal{U}}} \hat{\mathcal{J}}(\theta) = o(r). \tag{2.11}$$

Moreover, if $f \in L^\infty(0,1)$ and problem (2.4) has a solution θ_0 in $BV(0,1)$, then

$$0 \leq \min_{\theta \in \hat{\mathcal{U}}^r} \hat{\mathcal{J}}(\theta) - \min_{\theta \in \hat{\mathcal{U}}} \hat{\mathcal{J}}(\theta) \leq Cr^2. \tag{2.12}$$

Theorem 2.7 *Assuming $f \in L^1(0,1)$, problem (1.8) has a solution for every $r > 0$, and we have*

$$0 \leq \min_{\omega \in \mathcal{U}^r} \mathcal{J}(\omega) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega) \leq Cr^{\frac{1}{2}}. \tag{2.13}$$

Moreover, if for some integer $l \geq 1$, we have that f belongs to $W^{l,1}(0,1)$ and $F_1(x,s,\xi)$, $F_2(x,s,\xi)$ are independent of s and belong to $C^{l,1}_{loc}([0,1] \times \mathbb{R})$, then we have

$$0 \leq \min_{\omega \in \mathcal{U}^r} \mathcal{J}(\omega) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega) \leq Cr^{\frac{l+1}{l+2}}. \tag{2.14}$$

2.2 Optimality

We now give an example showing that the previous results are nearly optimal:

Proposition 2.8 *We consider problem (1.1) with $\alpha < \beta$, $f = 1$, $\kappa = 2/3$ and $\mathcal{J}$ given by*

$$\mathcal{J}(\omega) = -\alpha \int_\omega \left| \frac{du_\omega}{dx} \right|^2 dx - \beta \int_{(0,1)\setminus\omega} \left| \frac{du_\omega}{dx} \right|^2 dx. \tag{2.15}$$

For every $n \in \mathbb{N}$, we define $\mathcal{P}_n$ as the partition of $[0,1]$ given by

$$\mathcal{P}_n = \left\{ k10^{-n} : \ 0 \leq k \leq 10^n \right\}.$$

We define

$$\hat{\mathcal{U}}^n = \left\{ \theta \in \hat{\mathcal{U}} \ : \ \theta = \sum_{k=1}^{10^n} r_k \chi_{((k-1)10^{-n}, k10^{-n})}, \quad \text{with } r_k \in [0,1], \forall k \in \{1, \dots, 10^n\} \right\} \tag{2.16}$$

$$\mathcal{U}^n = \left\{ \omega \in \mathcal{U} : \exists I \subset \mathcal{P}_n \setminus \{1\} \text{ with } \omega = \bigcup_{k \in I} \left(k10^{-n}, (k+1)10^{-n} \right) \right\}. \tag{2.17}$$

Then, we have

$$\lim_{n \to \infty} \frac{\min_{\theta \in \hat{\mathcal{U}}^n} \hat{\mathcal{J}}(\theta) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega)}{10^{-2n}} = \frac{2(\beta - \alpha)}{27\alpha\beta} \tag{2.18}$$

$$\lim_{n \to \infty} \frac{\min_{\omega \in \mathcal{U}^n} \mathcal{J}(\omega) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega)}{10^{-n}} = \frac{\beta - \alpha}{54\alpha\beta}. \tag{2.19}$$

Remark 2.9 *This result shows that:*

- *Estimate (2.12) corresponding to the case in which the relaxed version of the problem is approximated is optimal.*
- *The estimate (2.14) for the case where the original problem is discretized is nearly optimal as well, in the sense that the upper bound can not be of the order of $o(r)$ as in (2.11).*

  However the question remains whether we can replace the second member of (2.14) by Cr.

The example considered in Proposition 2.8 is very particular. In this case problem (2.4) has the unique solution

$$\hat{\theta}_0 = \chi_{(0,1/3)\cup(2/3,1)} \tag{2.20}$$

(see the proof of Proposition 2.8 in Section 6). Since $\hat{\theta}_0$ is a characteristic function, we are in a case where problem (1.1) has a solution as well. Even in this case, as predicted by the theory, the error for the discretized relaxed problem (1.10) is a lot smaller than for the discretized unrelaxed one (1.8).

2.3 Direct versus relaxed discretization

By Theorems 2.6, 2.7 and Proposition 2.8 it is clear that in order to obtain an approximation of a solution of (2.4) it is better to use (1.10) than (1.8). Moreover, (1.10) is simpler to solve because the set of controls is a convex set while in (1.8) we are minimizing in a set of functions which only take the values 0 or 1. The unique advantage of (1.8) with respect to (1.10) is that it provides a physical solution and not a relaxed control.

The following proposition shows that this is not a great advantage because it is very simple to obtain a good unrelaxed control from a relaxed one. See Section 4 for its proof.

Proposition 2.10 *We assume $f \in L^\infty(0,1)$. Let $\mathcal{P}^r = \{y_k\}_{k=0}^{m_r}$, with $m_r \in \mathbb{N}$, be a partition of $[0,1]$ with r as in (2.8). Assume that,*

$$\theta = \sum_{k=1}^{m_r} t_k \chi_{(y_{k-1},y_k)} \in \hat{\mathcal{U}}^r,$$

with $t_k \in [0,1]$ for every $k \in \{1,\dots,m_r\}$. Taking

$$j_k = \left[\frac{y_k - y_{k-1}}{r^2}\right] + 1,\ s_k = \frac{y_k - y_{k-1}}{j_k}, \quad \forall\, k \in \{1,\dots,m_r\},$$

we define $\omega \subset (0,1)$ as

$$\omega = \bigcup_{k=1}^{m_r} \bigcup_{i=1}^{j_k} (y_{k-1} + (i-1)s_k, y_{k-1} + (i-1+t_k)s_k). \tag{2.21}$$

Then, we have

$$\left|\hat{\jmath}(\theta) - \jmath(\omega)\right| = \left|\hat{\jmath}(\theta) - \hat{\jmath}(\chi_\omega)\right| \le Cr^2, \tag{2.22}$$

whatever the functional $\hat{\jmath}$ is within the class of those considered in the general results of Section 2.1.

2.4 Finite element approximation

So far we have focused on the discretization of the admissible set of controls. However, a full discretization of the minimization problem (1.1) requires also the numerical approximation of (2.1) and the cost functional (2.2). The aim of this section is to analyze this fully discrete problem in order to see if the finite-element approximation of the relaxed formulation provides better approximations than the finite-element approximation of the direct optimization problem.

We first consider the finite element approximation of the non-relaxed problem. For $h > 0$, we introduce a second partition $\mathcal{P}^h = \{x_i\}_{i=0}^{n_h}$ of $[0,1]$, with

$$h = \max_{1\le i\le n_h} (x_i - x_{i-1}) \tag{2.23}$$

and we denote by W^h the space of finite elements

$$W^h = \{v \in C_0^0([0,1]) \,:\, v \text{ is affine on } (x_{i-1}, x_i),\ 1 \le i \le n_h\}. \tag{2.24}$$

Then, for every $\omega \in \mathcal{U}$ we introduce the finite-element approximation u_ω^h of u as the solution of the following finite-dimensional variational problem:

$$\begin{cases} u_\omega^h \in W^h \\ \displaystyle\int_0^1 (\alpha\chi_\omega + \beta(1-\chi_\omega))\frac{du_\omega^h}{dx}\frac{dv}{dx}dx = \int_0^1 fv dx, \qquad \forall v \in W^h. \end{cases} \tag{2.25}$$

We also set

$$\mathcal{J}^h(\omega) = \int_\omega F_1\left(x, u_\omega^h, \frac{du_\omega^h}{dx}\right)dx + \int_{(0,1)\setminus\omega} F_2\left(x, u_\omega^h, \frac{du_\omega^h}{dx}\right)dx, \quad \forall \omega \in \mathcal{U}. \tag{2.26}$$

Once we have introduced a natural finite-element approximation to evaluate the cost functional we can state the fully discrete optimization problem defined by (1.12).

We now introduce the finite-element approximation of the relaxed formulation. For every $\theta \in \hat{\mathcal{U}}^h$, defined by (2.9) we introduce the finite-element approximation $\tilde{u}_\theta$ as the solution of the following finite-dimensional variational problem:

$$\begin{cases} \tilde{u}_\theta \in W^h \\ \displaystyle\int_0^1 M_\theta \frac{d\tilde{u}_\theta}{dx}\frac{dv}{dx}dx = \int_0^1 fv dx, \quad \forall v \in W^h. \end{cases} \tag{2.27}$$

We also set

$$\hat{\mathcal{J}}^h(\theta) = \int_0^1 \left(\theta F_1\left(x, \tilde{u}_\theta, \frac{M_\theta}{\alpha}\frac{d\tilde{u}_\theta}{dx}\right) + (1-\theta)F_2\left(x, \tilde{u}_\theta\frac{M_\theta}{\beta}\frac{d\tilde{u}_\theta}{dx}\right)\right)dx \tag{2.28}$$

for the relaxed functional evaluated on the finite-element approximation. Note that, in the particular case $\theta = \chi_\omega$, we have

$$\mathcal{J}^h(\omega) = \hat{\mathcal{J}}^h(\chi_\omega). \tag{2.29}$$

Remark that $\hat{\mathcal{J}}^h$ is a discretized version of the relaxed functional $\hat{\mathcal{J}}$ and $\mathcal{J}^h$ is a discretized version of the unrelaxed functional $\mathcal{J}$.

The following result is the key ingredient in our convergence results:

Lemma 2.11 *Assume that $r \ge h$ and $\mathcal{P}^h$ is a refinement of $\mathcal{P}^r$. For every $f \in L^1(0,1)$, there exists a constant $C > 0$ such that*

$$\left|\hat{\mathcal{J}}(\theta) - \hat{\mathcal{J}}^h(\theta)\right| \le Ch, \quad \forall\, \theta \in \hat{\mathcal{U}}^r, \tag{2.30}$$

for all functionals and finite element approximations as above.

The condition $r \geq h$ in Lemma 2.11 is necessary, in general, as we show below.

By Theorem 2.6, Theorem 2.7, Proposition 2.10 and Lemma 2.11 we have the following two corollaries providing a numerical approximation of the control problem. Corollary 2.12 concerns with the discretization of the relaxed problem (2.4) while Corollary 2.13 concerns with the discretization of the original problem (1.1).

Corollary 2.12 *Assume $f \in L^\infty(0,1)$ and suppose that there exists an optimal control θ of the relaxed problem (2.4) which is of bounded variation in $(0,1)$.*

For $h > 0$, we denote $r = \sqrt{h}$ and we consider two partitions $\mathcal{P}^r = \{y_i\}_{i=1}^{m_r}$, $\mathcal{P}^h = \{x_i\}_{i=1}^{n_h}$ of $[0,1]$, with $\mathcal{P}^h$ a refinement of $\mathcal{P}^r$ fulfilling (2.23) and (2.8).

Defining $\hat{\mathcal{U}}^r$ by (2.9), we consider the full discrete problem (1.14) with $\hat{\mathcal{J}}^h$ defined by (2.28), which has a solution.

Then, every solution θ_0 of (1.14) satisfies

$$0 \leq \mathcal{J}(\omega_0) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega) \leq Ch, \tag{2.31}$$

where the unrelaxed control $\omega_0 \in \mathcal{U}$ is defined from θ_0 by the mechanism (2.21).

Corollary 2.13 *For $f \in L^\infty(0,1)$ and $h > 0$ we consider a partition $\mathcal{P}^h = \{x_i\}_{i=1}^{n_h}$ of $[0,1]$, satisfying (2.23). We consider the control problem*

$$\min_{\omega \in \mathcal{U}^h} \mathcal{J}^h(\omega), \tag{2.32}$$

with $\mathcal{U}^h$ defined by (2.10) (with $h \leq r$ and $\mathcal{P}^h$ a refinement of $\mathcal{P}^r$) and $\mathcal{J}^h$ defined by (2.29), which has a solution.

Then, every solution ω_0 of (2.32) satisfies

$$0 \leq \mathcal{J}(\omega_0) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega) \leq Cr^{\frac{1}{2}}. \tag{2.33}$$

Moreover, if for some nonnegative integer, we have that f belongs to $W^{l,1}(0,1)$ and $F_1(x,s,\xi)$, $F_2(x,s,\xi)$ are independent of s and belong to $C^{l,1}_{loc}([0,1]\times\mathbb{R})$, then we have

$$0 \leq \mathcal{J}(\omega_0) - \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega) \leq Cr^{\frac{l+1}{l+2}}. \tag{2.34}$$

Remark 2.14 *Solving the corresponding finite-element control problems, Corollaries 2.12 and 2.13 provide a physical control $\omega_0 \in \mathcal{U}$ such that $\mathcal{J}(\omega_0)$ is close to the infimum of $\mathcal{J}$.*

From a computational point of view, the discretization considered in Corollary 2.12 is better than the one considered in Corollary 2.13 not only because the error is slightly better but also because in Corollary 2.12 the set of controls is convex and so the discretized problem (1.14) is simpler to solve. Moreover, the elements of the partition where the controls are constant are a lot larger in Corollary 2.12 than in Corollary 2.13. This reduces considerably the computational cost.

In Corollary 2.12 we have supposed f in $L^\infty(0,1)$ and the existence of an optimal control of bounded variation. If this is not satisfed, then taking in Corollary 2.12 $r = h$ we still have an estimate of order h in (2.31) thanks to (2.11).

2.5 The case $r < h$

In the convergence results of the previous section we assumed $r \geq h$. Here we give two examples which show that if $r < h$ some undesirable situations may appear. To fix ideas we focus on the particular case $r = h/2$. The key point is the following lemma which establishes that the result in Lemma 2.11 may fail in this situation.

Lemma 2.15 *Let $h = 1/k$ with $k \in \mathbb{N}$, let $\mathcal{P}^h = \{x_j\}_{j=0}^k$, $\mathcal{P}^{h/2} = \{y_l\}_{l=0}^{2k}$ be the uniform partitions of $[0,1]$ constituted by $x_j = jh$, $j = 0, 1, \ldots, k$ and $y_l = lh/2$, $l = 0, 1, \ldots, 2k$ and let*

$$\omega^{h/2} = \bigcup_{j=0}^{k-1} \left(\frac{j}{k}, \frac{j}{k} + \frac{1}{2k} \right) \in \mathcal{U}^{h/2}. \tag{2.35}$$

Then,

$$\lim_{h\to 0} \mathcal{J}(\omega^{h/2}) = \hat{\mathcal{J}}(\theta_0), \qquad \lim_{h\to 0} \mathcal{J}^h(\omega^{h/2}) = \hat{\mathcal{J}}(\theta_m), \tag{2.36}$$

where $\theta_0 = 1/2$ and $\theta_m = \alpha/(\alpha+\beta)$. In particular, if $\hat{\mathcal{J}}(\theta_0) \neq \hat{\mathcal{J}}(\theta_m)$ then (2.30) will not hold.

We prove this lemma in section 7 below.

Based on this result we show now two examples which exhibit the lack of convergence of the fully discrete optimization problems.

Example 1. This example shows how minimizing sequences of the continuous optimization problem can be far from being discrete optima when $h << 1$. In particular, this means that any numerical algorithm able to solve the discrete optimization problem for h small will not provide such minimizing sequences of the continuous problem.

We consider the minimization problem (1.1), with $f = 1$, $\kappa = 1/2$, and the functional

$$\mathcal{J}(\omega) = \int_0^1 |u(x) - u^*(x)|^2 dx, \tag{2.37}$$

where $u^*(x) = (x - x^2)/2a^*$ and $a^* = M(1/2)$ is the harmonic mean of α and β with proportion $1/2$. According to Theorem 2.2, a relaxation of this problem is given by (2.4). Note that the relaxed problem has a unique minimizer corresponding to

$$\theta_{\min} = 1/2,$$

since, in this case, the solution $u_{\theta_{\min}}$ of (2.7) coincides with u^* and $\hat{\mathcal{J}}(\theta_{\min}) = 0$. Thus, this is a case where the original problem (1.1) does not have a minimizer in $\mathcal{U}$.

Let us consider now the discretization of (1.1) given by (1.12), associated to the uniform partition $\mathcal{P}^h = \{y_j\}_{j=0}^{m_h}$ where $y_j = jh$ and $m_h = 1/h \in \mathbb{N}$.

From Corollaries 2.12 and 2.13 we see that

$$\lim_{h\to 0} \min_{\omega\in\mathcal{U}^h} \mathcal{J}^h(\omega) = \lim_{h\to 0} \min_{\theta\in\hat{\mathcal{U}}^h} \hat{\mathcal{J}}^h(\theta) = \inf_{\omega\in\mathcal{U}} \mathcal{J}(\omega) = 0.$$

Moreover, minimizing sequences of the continuous problem and minimizers of the discrete functionals as $h \to 0$ are related, due to Lemma 2.11. More precisely, in the context of the non-relaxed problem, minimizers of $\mathcal{J}^h$ in $\mathcal{U}^h$ constitute a minimizing sequence for the continuous problem as $h \to 0$. On the other hand, any minimizing sequence of the continuous problem ω_m^h constituted by elements in $\mathcal{U}^h$ as $h \to 0$, i.e. $\omega_m^h \in \mathcal{U}^h$, is close to a minimizer of $\mathcal{J}^h$ in $\mathcal{U}^h$ in the sense that

$$\lim_{h\to 0} (\mathcal{J}^h(\omega_m^h) - \min_{\omega\in\mathcal{U}^h} \mathcal{J}^h(\omega)) = 0.$$

Let us consider now the sequence $\omega^{h/2} \in \mathcal{U}^{h/2}$ defined in (2.35). It is easy to see that it constitutes a minimizing sequence as $h \to 0$. In fact, as stated in Lemma 2.15, the solution of (2.1) with $\omega = \omega^{h/2}$, that we write $u^{h/2}(x)$, satisfies

$$u^*(x) = \lim_{h\to 0} u^{h/2}(x),$$

and therefore $\mathcal{J}(\omega^{h/2}) \to 0$ as $h \to 0$.

A rather natural conjecture is to think that $\mathcal{J}^h(\omega^{h/2})$ should be close to $\inf_{\omega \in \mathcal{U}^{h/2}} \mathcal{J}^h(\omega)$ as $h \to 0$. We see that this is not the case.

First of all, note that, as stated in Lemma 2.15,

$$\lim_{h\to 0} \mathcal{J}^h(\omega^{h/2}) = \hat{\mathcal{J}}(\theta_m) = \hat{\mathcal{J}}(\frac{\alpha}{\alpha+\beta}) > 0.$$

On the other hand, we remark that $\lim_{h\to 0} \inf_{\omega \in \mathcal{U}^{h/2}} \mathcal{J}^h(\omega) = 0$ since

$$0 \leq \inf_{\omega \in \mathcal{U}^{h/2}} \mathcal{J}^h(\omega) \leq \inf_{\omega \in \mathcal{U}^h} \mathcal{J}^h(\omega),$$

and the right hand side converges to zero, as $h \to 0$, as we have seen before. This shows that the discrete method corresponding to take $r = h/2$ converges in this case. Let us show in the next example that this does not always holds.

Example 2. This example shows that the value of the discrete functional at discrete optima may not converge to the infimum of the continuous functional, as $h \to 0$. For $\alpha > \beta > 0$ and $\kappa = 1/2$, we consider problem (1.1), for

$$\mathcal{J}(\omega) = \int_0^1 |u_\omega(x) - u^*(x)|^2 dx,$$

with $u^*(x) = (x^2 - x)/(\alpha + \beta)$ the solution of

$$\begin{cases} -\dfrac{\alpha+\beta}{2} \dfrac{d^2u^*}{dx} = 1 \ \text{ in } (0,1) \\ u^*(0) = u^*(1) = 0. \end{cases}$$

Proposition 2.16 *For $h = 1/k$, with $k \in \mathbb{N}$, we take $\mathcal{P}^h = \{x_j\}_{j=0}^k$ and $\mathcal{P}^{h/2} = \{y_l\}_{l=0}^{2k}$ as the uniform partitions of $[0,1]$ constituted by $x_j = jh$, $j = 0, 1, \ldots, k$ and $y_l = lh/2$, $l = 0, 1, \ldots, 2k$. Then, we have,*

$$0 = \lim_{h\to 0} \min_{\theta \in \hat{\mathcal{U}}^{h/2}} \hat{\mathcal{J}}^h(\theta) = \lim_{h\to 0} \min_{\omega \in \mathcal{U}^{h/2}} \mathcal{J}^h(\omega) < \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega). \tag{2.38}$$

Proof. For $k \in \mathbb{N}$, we take $\omega^{h/2} \in \mathcal{U}^{h/2}$ as in (2.35). Then, we observe that the solution u^h of

$$\begin{cases} u^h \in W^h \\ \displaystyle\int_0^1 (\alpha \chi_{\omega^{h/2}} + \beta(1 - \chi_{\omega^{h/2}})) \frac{du^h}{dx} \frac{dv}{dx} = \int_0^1 v\, dx\,, \qquad \forall\, v \in W^h, \end{cases}$$

agrees with the solution $u^{*,h}$ of

$$\begin{cases} u^{*,h} \in W^h \\ \dfrac{\alpha+\beta}{2} \displaystyle\int_0^1 \frac{du^{*,h}}{dx} \frac{dv}{dx} = \int_0^1 v\, dx\,, \qquad \forall\, v \in W^h. \end{cases}$$

Then, by the classical estimate for the solutions of elliptic equations via finite elements, we know

$$\|u^{*,h} - u\|_{H^1(0,1)} \leq Ch,$$

which proves

$$0 \leq \min_{\theta \in \hat{\mathcal{U}}^{h/2}} \hat{\mathcal{J}}^h(\theta) \leq \min_{\omega \in \mathcal{U}^{h/2}} \mathcal{J}^h(\omega) \leq Ch^2.$$

This gives the equalities in (2.38). However, let us prove by contradiction that

$$0 < \inf_{\omega \in \mathcal{U}} \mathcal{J}(\omega).$$

If not, by Theorem 2.2 there exists $\theta \in \hat{\mathcal{U}}$, such that u^* satisfies

$$-\frac{d}{dx}\left(M(\theta)\frac{du^*}{dx}\right) = 1 \ \text{ in } (0,1),$$

which implies that there exists a constant c such that

$$\left(M(\theta) - \frac{\alpha+\beta}{2}\right)\frac{du^*}{dx} = c.$$

Taking into account that $M(\theta)\frac{du^*}{dx}$ is a continuous function and $\frac{du^*}{dx}(1/2) = 0$, we obtain that $c = 0$ and then that $M(\theta) = \frac{\alpha+\beta}{2}$ for a.e. θ, i.e.

$$\theta = \frac{\alpha}{\alpha+\beta}, \ \text{ a.e. in } (0,1).$$

However, since we are assuming that $\alpha > \beta$, this θ satisfies

$$\int_0^1 \theta dx = \frac{\alpha}{\alpha+\beta} > \frac{1}{2},$$

in contradiction with the volume restriction. □

3 Proof of the relaxation result

This section is devoted to prove Theorem 2.2 which characterizes the relaxation of problem (1.1). To do it, we use the following lemma.

Lemma 3.1 *The functional $\hat{\mathcal{J}} : \hat{\mathcal{U}} \subset L^\infty(0,1) \to \mathbb{R}$ is sequentially continuous for the $*$-weak topology of $L^\infty(0,1)$.*

Proof. Given a sequence $\theta_n \in \hat{\mathcal{U}}$ which converges weakly-$*$ in $L^\infty(0,1)$ to a function $\theta \in \hat{\mathcal{U}}$, we have to see that $\hat{\mathcal{J}}(\theta_n)$ converges to $\hat{\mathcal{J}}(\theta)$. For a such sequence θ_n, we observe that the corresponding solution u_{θ_n} of (2.7) is given by

$$u_{\theta_n}(x) = -\int_0^x \frac{F(t) - c_n}{M_{\theta_n}(t)} dt = -\int_0^x \big(F(t) - c_n\big)\frac{\alpha(1-\theta_n(t)) + \beta\theta_n(t)}{\alpha\beta} dt,$$

with F a primitive of f in $(0,1)$ and

$$c_n = \left(\int_0^1 \frac{dt}{M_{\theta_n}(t)}\right)^{-1}\left(\int_0^1 \frac{F(t)}{M_{\theta_n}(t)}\, dt\right).$$

Therefore, it is immediate to show that

$$\|u_{\theta_n}\|_{W^{1,\infty}(0,1)} \le C, \quad u_{\theta_n} \to u_\theta \ \text{ in } C^0([0,1]), \quad M_{\theta_n}\frac{du_n}{dx} - M_\theta\frac{du_\theta}{dx} \to 0 \ \text{ in } C^0([0,1]),$$

with u_θ the unique solution of (2.7). Then, by (2.3) we obtain

$$\lim_{n\to\infty} \hat{\mathcal{J}}(\theta_n) = \lim_{n\to\infty} \int_0^1 \left(\theta_n F_1\left(x, u_{\theta_n}, \frac{M_{\theta_n}}{\alpha}\frac{du_{\theta_n}}{dx}\right) + (1-\theta_n)F_2\left(x, u_{\theta_n}, \frac{M_{\theta_n}}{\beta}\frac{du_{\theta_n}}{dx}\right)\right) dx$$
$$= \int_0^1 \left(\theta F_1\left(x, u_\theta, \frac{M_\theta}{\alpha}\frac{du_\theta}{dx}\right) + (1-\theta)F_2\left(x, u_\theta, \frac{M_\theta}{\beta}\frac{du_\theta}{dx}\right)\right) dx = \hat{\mathcal{J}}(\theta).$$

□

Proof of Theorem 2.2. Taking into account that the space of controls $\hat{\mathcal{U}}$ given by (2.5) is sequentially compact in the $*$-weak topology of $L^\infty(0,1)$, from Lemma 3.1 we deduce that problem (2.4) has at least a solution. On the other hand, by Remark 2.4 it is clear that

$$\inf_{\omega\in\mathcal{U}} \mathcal{J}(\omega) = \inf_{\chi_\omega\in\hat{\mathcal{U}}} \hat{\mathcal{J}}(\chi_\omega) \geq \min_{\theta\in\hat{\mathcal{U}}} \hat{\mathcal{J}}(\theta).$$

Therefore, in order to check that problem (2.4) is a relaxation of (1.1), it is enough to prove that for every $\theta \in \hat{\mathcal{U}}$, there exists a sequence ω_n in $\mathcal{U}$ such that

$$\chi_{\omega_n} \overset{*}{\rightharpoonup} \theta \ \text{ in } L^\infty(0,1) \tag{3.1}$$

$$\mathcal{J}(\omega_n) \to \hat{\mathcal{J}}(\theta). \tag{3.2}$$

The existence of this sequence ω_n is well known (for example it is a consequence of Lemma 5.1 below), while by the continuity property of $\hat{\mathcal{J}}$ proved in Step 1, (3.2) is a consequence of (3.1). So, the proof of Theorem 2.2 is complete. □

4 Proof of the convergence estimates for the discretized relaxed control problem

In this section we prove Theorem 2.6 referred to the convergence of the discretization of problem (2.4) given by (1.10). Note that we are discretizing the controls but not the state equation. We also give the proof of Proposition 2.10 which permits to obtain a physical control from a relaxed one.

Along this section, we consider a partition $\mathcal{P}^r = \{y_k\}_{k=0}^{m_r}$, with $m_r \in \mathbb{N}$, satisfying (2.8). The space $\hat{\mathcal{U}}^r$ is defined by (2.9).

In order to show Theorem 2.6 we will use the operator Π^r defined by

Definition 4.1 *We define the projection operator* $\Pi^r : L^1(0,1) \longrightarrow \hat{\mathcal{U}}^r$ *by*

$$\Pi^r\psi = \sum_{k=1}^{m_r} \fint_{y_{k-1}}^{y_k} \psi\, ds\ \chi_{(y_{k-1},y_k)}, \quad \forall \psi \in L^1(0,1). \tag{4.1}$$

The following Lemma estimates the difference $\Pi^r\theta - \theta$ when r tends to zero.

Lemma 4.2 *Let* θ *be in* $L^\infty(0,1;[0,1])$. *Then, for every* $\varphi \in W^{1,1}(0,1)$, *it holds*

$$\int_0^1 (\theta - \Pi^r\theta)\,\varphi\, dx = o(r) \tag{4.2}$$

$$\int_0^1 \left| \int_0^x (\theta(t) - \Pi^r\theta(t))\,\varphi(t)\, dt \right| dx = o(r). \tag{4.3}$$

Moreover, if θ *is in* $BV(0,1)$, *and* φ *in* $W^{1,\infty}(0,1)$, *we have the following improvement of the previous estimates*

$$\left| \int_0^1 (\theta - \Pi^r\theta)\,\varphi\, dx \right| \leq C \left\| \frac{d\varphi}{dx} \right\|_{L^\infty(0,1)} r^2 \tag{4.4}$$

$$\int_0^1 \left| \int_0^x (\theta(t) - \Pi^r\theta(t))\,\varphi(t)\, dt \right| dx \leq C \|\varphi\|_{W^{1,\infty}(0,1)} r^2. \tag{4.5}$$

Proof. We take $\varphi \in W^{1,1}(0,1)$, for a given $x \in [0,1]$, we consider y_j defined by

$$y_j = \sup\{y_k : \ y_k \le x, \ 0 \le k \le m_r\}.$$

Then, using the inequality

$$\left| \varphi(t) - \fint_{y_{k-1}}^{y_k} \varphi \, ds \right| \le \left\| \frac{d\varphi}{dt} \right\|_{L^1(y_{k-1},y_k)}, \quad \forall \, t \in [y_{k-1}, y_k]$$

we have

$$\begin{aligned}
&\left| \int_0^x (\theta - \Pi^r \theta) \, \varphi \, dt \right| \\
&= \sum_{k=1}^{j} \int_{y_{k-1}}^{y_k} \left(\theta - \fint_{y_{k-1}}^{y_k} \theta \, ds \right) \varphi \, dt + \int_{y_j}^{x} \left(\theta - \fint_{y_j}^{y_{j+1}} \theta \, ds \right) \varphi \, dt \\
&= \sum_{k=1}^{j} \int_{y_{k-1}}^{y_k} \left(\theta - \fint_{y_{k-1}}^{y_k} \theta \, ds \right) \left(\varphi - \fint_{y_{k-1}}^{y_k} \varphi \, ds \right) dt + \int_{y_j}^{x} \left(\theta - \fint_{y_j}^{y_{j+1}} \theta \, ds \right) \varphi \, dt \\
&\le \sum_{k=1}^{j} \left\| \frac{d\varphi}{dx} \right\|_{L^1(y_{k-1},y_k)} \|\theta - \Pi^r \theta\|_{L^1(y_{k-1},y_k)} + \|\varphi\|_{L^\infty(0,1)} \|\theta - \Pi^r \theta\|_{L^1(y_j,x)}.
\end{aligned} \tag{4.6}$$

Integrating this inequality in $(0,1)$, we get

$$\begin{aligned}
&\int_0^1 \left| \int_0^x (\theta(t) - \Pi^r \theta(t)) \, \varphi(t) \, dt \right| dx \\
&\le \sum_{k=1}^{m_r} \left\| \frac{d\varphi}{dx} \right\|_{L^1(y_{k-1},y_k)} \|\theta - \Pi^r \theta\|_{L^1(y_{k-1},y_k)} + \|\varphi\|_{L^\infty(0,1)} \sum_{j=0}^{m_r - 1} \int_{y_j}^{y_{j+1}} \|\theta - \Pi^r \theta\|_{L^1(y_j,x)} dx \\
&\le \sum_{k=1}^{m_r} \left\| \frac{d\varphi}{dx} \right\|_{L^1(y_{k-1},y_k)} \|\theta - \Pi^r \theta\|_{L^1(y_{k-1},y_k)} + \|\varphi\|_{L^\infty(0,1)} \|\theta - \Pi^r \theta\|_{L^1(0,1)} \, r.
\end{aligned} \tag{4.7}$$

If φ belongs to $W^{1,\infty}(0,1)$ and θ belongs to $BV(0,1)$, using in (4.7)

$$\left\| \frac{d\varphi}{dx} \right\|_{L^1(y_{k-1},y_k)} \le \left\| \frac{d\varphi}{dx} \right\|_{L^\infty(0,1)} r, \quad \|\theta - \Pi^r \theta\|_{L^1(0,1)} \le V_{(0,1)}(\theta) \, r, \tag{4.8}$$

we deduce (4.5).

Inequality (4.4) is a consequence of (4.6) with $x = 1 = y_j$ and (4.8).

In order to show (4.2) and (4.3) we now take a sequence φ_n in $W^{1,\infty}(0,1)$ which converges to φ in $W^{1,1}(0,1)$ and a sequence θ_n in $BV(0,1)$, with $0 \le \theta_n \le 1$ in $(0,1)$, which converges to θ in $L^1(0,1)$. Then, we estimate the right-hand side of (4.7) as follows

$$\begin{aligned}
&\sum_{k=1}^{m_r} \left\| \frac{d\varphi}{dx} \right\|_{L^1(y_{k-1},y_k)} \|\theta - \Pi^r \theta\|_{L^1(y_{k-1},y_k)} + \|\varphi\|_{L^\infty(0,1)} \|\theta - \Pi^r \theta\|_{L^1(0,1)} \, r \\
&\le 2 \left\| \frac{d(\varphi - \varphi_n)}{dx} \right\|_{L^1(0,1)} r + \|\varphi\|_{L^\infty(0,1)} \|\theta - \theta_n - \Pi^r(\theta - \theta_n)\|_{L^1(0,1)} \, r \\
&\quad + \sum_{k=1}^{m_r} \left\| \frac{d\varphi_n}{dx} \right\|_{L^1(y_{k-1},y_k)} \|\theta - \Pi^r \theta\|_{L^1(y_{k-1},y_k)} \ + \|\varphi\|_{L^\infty(0,1)} \|\theta_n - \Pi^r \theta_n\|_{L^1(0,1)} \, r \\
&\le 2 \left\| \frac{d(\varphi - \varphi_n)}{dx} \right\|_{L^1(0,1)} r + \|\varphi\|_{L^\infty(0,1)} \|\theta - \theta_n - \Pi^r(\theta - \theta_n)\|_{L^1(0,1)} \, r \\
&\quad + \left(\left\| \frac{d\varphi_n}{dx} \right\|_{L^\infty(0,1)} V_{(0,1)}(\theta) + \|\varphi\|_{L^\infty(0,1)} V_{(0,1)}(\theta_n) \right) r^2.
\end{aligned}$$

Dividing this inequality by r and passing to the limit first when r tends to zero and then when n tends to infinity, we deduce (4.3). The proof of (4.2) can be obtained reasoning in a similar way with (4.6). □

For $\theta \in L^\infty(0,1;[0,1])$, the following lemma estimates the difference between the solution of (2.7) and the solution of the analogous problem when θ is replaced by $\Pi^r\theta$.

Lemma 4.3 *Assume $f \in L^1(0,1)$. For $\theta \in L^\infty(0,1;[0,1])$, we consider $\theta^r = \Pi^r\theta$. Then, the solutions u_θ and u_{θ^r} of (2.7) for θ and θ^r respectively, satisfy*

$$\|u_\theta - u_{\theta^r}\|_{L^1(0,1)} \le o(r) \tag{4.9}$$

$$\left\| M_\theta \frac{du_\theta}{dx} - M_{\theta^r} \frac{du_{\theta^r}}{dx} \right\|_{L^\infty(0,1)} \le o(r). \tag{4.10}$$

If f is in $L^\infty(0,1)$ and θ is in $BV(0,1)$, then in (4.9) and (4.10) we can take

$$o(r) = C\, V_{(0,1)}(\theta)\, r^2.$$

Proof. The functions u_θ and u_{θ^r} are given by

$$u_\theta(x) = -\int_0^x \frac{g}{M_\theta} ds + c \int_0^x \frac{1}{M_\theta} ds \quad \text{for a.e. } x \in (0,1) \tag{4.11}$$

$$u_{\theta^r}(x) = -\int_0^x \frac{g}{M_{\theta^r}} ds + c_r \int_0^x \frac{1}{M_{\theta^r}} ds \quad \text{for a.e. } x \in (0,1) \tag{4.12}$$

with g a primitive of f and $c, c_r \in \mathbb{R}$ defined by

$$c = \left(\int_0^1 \frac{1}{M_\theta} dx \right)^{-1} \int_0^1 \frac{g}{M_\theta}\, dx, \qquad c_r = \left(\int_0^1 \frac{1}{M_{\theta^r}} dx \right)^{-1} \int_0^1 \frac{g}{M_{\theta^r}}\, dx. \tag{4.13}$$

Using these expressions and taking into account that

$$\min\{\alpha,\beta\} \le M_\theta, M_{\theta^r} \le \max\{\alpha,\beta\},$$

we easily deduce

$$\begin{aligned} \|u_\theta - u_{\theta^r}\|_{L^1(0,1)} \le C \Bigg(& \left| \int_0^1 (\theta - \theta^r)\, g\, dx \right| + \left| \int_0^1 (\theta - \theta^r)\ dx \right| \\ & + \int_0^1 \left| \int_0^x (\theta(t) - \theta^r(t))\, g(t)\, dt \right| dx + \int_0^1 \left| \int_0^x (\theta(t) - \theta^r(t))\ dt \right| dx \Bigg) \end{aligned}$$

and

$$\left\| M_\theta \frac{du_\theta}{dx} - M_{\theta^r} \frac{du_{\theta^r}}{dx} \right\|_{L^\infty(0,1)} \le C \left(\left| \int_0^1 (\theta - \theta^r)\, g\, dx \right| + \left| \int_0^1 (\theta - \theta^r)\ dx \right| \right).$$

Lemma 4.3 is then a simple consequence of Lemma 4.2. □

We are now in position to prove

Proof of Theorem 2.6. The existence of solution for problem (1.10) is a simple consequence of the compactness of (2.9) in $L^1(0,1)$.

On the other hand, using that F_1 and F_2 are locally Lipschitz, and that the functions u_θ, u_{θ^r} defined as in Lemma 4.2 are bounded in $W^{1,\infty}(0,1)$ independently of r, we have

$$\begin{aligned} & |\hat{\mathcal{J}}(\theta) - \hat{\mathcal{J}}(\theta^r)| \\ & \le \left| \int_0^1 F_1\left(x, u_\theta, \frac{M_\theta}{\alpha}\frac{du_\theta}{dx}\right)(\theta - \theta^r)\, dx \right| + \left| \int_0^1 F_2\left(x, u_\theta, \frac{M_\theta}{\alpha}\frac{du_\theta}{dx}\right)(\theta - \theta^r)\, dx \right| \\ & \quad + C \int_0^1 \left(|u_\theta - u_{\theta^r}| + \left| M_\theta \frac{du_\theta}{dx} - M_{\theta^r} \frac{du_{\theta^r}}{dx} \right| \right) dx. \end{aligned}$$

Thanks to Lemma 4.3, we then deduce (2.11) and (2.12). □

To finish this section, we now give the proof of Proposition 2.10.

Proof of Proposition 2.10. Reasoning as in the proof of Theorem 2.6, we have that the result is an immediate consequence of the following lemma, which is similar to Lemma 4.2. □

Lemma 4.4 *Assume θ and ω as in the statement of Proposition 2.10, then for every $\varphi \in W^{1,\infty}(0,1)$, it holds*

$$\left| \int_0^1 (\theta - \chi_\omega)\, \varphi \, dx \right| \leq \left\| \frac{d\varphi}{dx} \right\|_{L^\infty(0,1)} r^2 \tag{4.14}$$

$$\int_0^1 \left| \int_0^x (\theta(t) - \chi_\omega(t))\, \varphi(t)\, dt \right| dx \leq \|\varphi\|_{W^{1,\infty}(0,1)}\, r^2. \tag{4.15}$$

Proof. Since in each interval $[y_{k-1} + (i-1)s_k, y_k + is_k]$, with $1 \leq k \leq m_r$, $1 \leq i \leq j_k$ the functions θ and χ_ω have the same integral, we can reason as in the proof of (4.6) to deduce that for every $x \in [0,1]$, we have

$$\left| \int_0^x (\theta - \chi_\omega)\, \varphi \, dt \right| \leq \left\| \frac{d\varphi}{dx} \right\|_{L^\infty(0,1)} \|\theta - \chi_\omega\|_{L^1(0,1)}\, r^2 + \|\varphi\|_{L^\infty(0,1)} \|\theta - \chi_\omega\|_{L^1(I)}\, r, \tag{4.16}$$

where I is an interval of the form $[y_{k-1} + (i-1)s_k, y_k + is_k]$ containing x. Taking $x = 1$ we get (4.14). On the other hand, since θ and χ_ω belong to $L^\infty(0,1;[0,1])$, inequality (4.16) implies

$$\left| \int_0^x (\theta - \chi_\omega)\, \varphi \, dt \right| \leq r^2\, \|\varphi\|_{W^{1,\infty}(0,1)}\,,$$

for every $x \in [0,1]$. This inequality immediately proves (4.15). □

5 Proof of the convergence estimates for the discretized unrelaxed control problem

Let us now prove Theorem 2.7. As for Theorem 2.6, we will need some preliminary lemmas.

Lemma 5.1 *We consider $\theta \in L^\infty(0,1)$ and $l \in \mathbb{N}$, then, there exists $\omega \subset (0,1)$ measurable such that*

$$\int_0^1 t^j \theta(t)\, dt = \int_\omega t^j dt, \quad \forall\, j \in \{0, \cdots, l\}. \tag{5.1}$$

Moreover ω can be chosen in the following way:

If $l = 2n$, with $n \in \mathbb{N}$,

$$\omega = (0, b_0) \bigcup \left(\bigcup_{i=1}^m (a_i, b_i) \right),$$

where $m \leq n$ and $0 \leq b_0 < a_1 < b_1 < \cdots < a_m < b_m \leq 1$.

If $l = 2n+1$, with $n \in \mathbb{N}$,

$$\omega = \bigcup_{i=1}^m (a_i, b_i),$$

where $m \leq n+1$ and $0 \leq a_1 < b_1 < \cdots < a_m < b_m \leq 1$.

Proof. Let us prove the result in the case $l = 2n+1$, the other one being similar.

We define $D \subset L^1(0,1)$ as

$$D = \left\{ \phi \in L^1(0,1) : \ \phi = \sum_{i=1}^{m} \chi_{(a_i,b_i)}, \quad \text{with } m \leq n+1, \ 0 \leq a_1 < b_1 < \cdots < a_m < b_m \leq 1 \right\}$$

and $\Psi : D \to \mathbb{R}$ by

$$\Psi(\phi) = \sum_{j=0}^{2n+1} \left(\int_0^1 t^j (\theta(t) - \phi(t))\, dt \right)^2, \quad \forall\, \phi \in D.$$

Since D is compact in $L^1(0,1)$ and Ψ is continuous, we know that Ψ attains its minimum in some function

$$\phi = \sum_{i=1}^{m} \chi_{(a_i,b_i)} \in D.$$

Then, we define the polynomial P as

$$P(\lambda) = \sum_{j=0}^{2n+1} \left(\int_0^1 t^j (\theta(t) - \phi(t))\, dt \right) \lambda^j$$

We fix k, with $1 \leq k \leq m$. For $\varepsilon \in \mathbb{R}$, with $|\varepsilon|$ small, $\varepsilon > 0$ if $k = 1$ and $a_1 = 0$ the function

$$\phi_\varepsilon = \chi_{\cup_{i \neq k}(a_i,b_i)} + \chi_{(a_k+\varepsilon,b_k)}$$

belongs to D. Taking into account that

$$\Psi(\phi_\varepsilon) = \sum_{j=0}^{2n+1} \left(\int_0^1 t^j (\theta(t) - \phi(t))\, dt + \int_{a_k}^{a_k+\varepsilon} t^j\, dt \right)^2,$$

and that ϕ is a minimum point of Ψ, we can derive with respect to ε in $\Psi(\phi_\varepsilon)$ to obtain that

$$P(a_k) = 0 \text{ if } a_k \neq 0, \qquad P(a_1) \geq 0 \text{ if } a_1 = 0.$$

Analgously, we can prove

$$P(b_k) = 0 \text{ if } b_k \neq 1, \qquad P(b_m) \geq 0 \text{ if } b_m = 1.$$

If P has $2n+2$ zeros, then it is the zero polynomial and we obtain the conclusion of the lemma. So, we assume in the following that P has at most $2n+1$ zeros. By the above proved we deduce that

$$m = n+1, \ a_1 = 0 \ \text{ and/or } b_{n+1} = 1,$$

or

$$m < n+1.$$

Let us prove that in all these cases P satisfies

$$P(\lambda) \geq 0 \ \text{ in } \bigcup_{i=1}^{m} (a_i, b_i), \quad P(\lambda) \leq 0 \ \text{ in } (0,1) \setminus \bigcup_{i=1}^{m} (a_i, b_i) \tag{5.2}$$

i) Case $m = n+1$, $a_1 = 0$, $b_{n+1} = 1$. Since we are supposing that the number of zeros of P is strictly less than $2n+2$ and P vanishes in the $2n$ points a_k with $k = 2, \cdots n+1$, b_k with $k = 1, \cdots, n$ we have that P has $2n$ or $2n+1$ zeros in $[0,1]$. If the number of zeros

is $2n+1$, then using that $P(0), P(1) \geq 0$ we deduce that the other zero of P is in 0 or 1 and that P satisfies (5.2). If the number of zeros is $2n$, then we have $P(0), P(1) > 0$ and (5.2) is satisfied.

ii) Case $m = n+1$, $a_1 = 0$, $b_{n+1} < 1$. In this case we have that the $2n+1$ zeros of P are given by the points a_k with $k = 2, \cdots n+1$, b_k with $k = 1, \cdots, n+1$. Since $P(0) \geq 0$, we deduce (5.2).

iii) Case $m = n+1$, $a_1 > 0$, $b_{n+1} = 1$. It is similar to the case ii).

iv) Case $m < n+1$. In this case, we take a point $c \in (a_i, b_i)$ for some $i \in \{1, \cdots m\}$. Then for $\varepsilon > 0$, small enough, the function

$$\phi_\varepsilon = \phi - \chi_{(c-\varepsilon, c+\varepsilon)}$$

belongs to D. Using that

$$\Psi(\phi_\varepsilon) = \sum_{j=0}^{2n+1} \left(\int_0^1 t^j (\theta(t) - \phi(t))\, dt + \int_{c-\varepsilon}^{c+\varepsilon} t^j\, dt \right)^2,$$

and deriving with respect to ε we deduce that

$$P(c) \geq 0, \quad \forall\, c \in \bigcup_{i=1}^{m} (a_i, b_i).$$

Analogously, if $c \in (0,1) \setminus \bigcup_{i=1}^m [a_i, b_i]$, taking

$$\phi_\varepsilon = \phi + \chi_{(c-\varepsilon, c+\varepsilon)},$$

we deduce that

$$P(c) \leq 0, \quad \forall\, c \in (0,1) \setminus \bigcup_{i=1}^{m} [a_i, b_i].$$

Thus, (5.2) is also proved in this case.

To finish, let us prove that (5.2) implies the conclusion of the Lemma. For this purpose, we just write

$$\begin{aligned} \sum_{j=0}^{2n+1} \left(\int_0^1 t^j (\theta(t) - \phi(t))\, dt \right)^2 &= \int_0^1 \left(\sum_{j=0}^{2n+1} \int_0^1 t^j (\theta(t) - \phi(t))\, dt\, s^j \right) (\theta(s) - \phi(s))\, ds \\ &= \int_0^1 P(s)(\theta(s) - \phi(s))\, ds, \end{aligned} \tag{5.3}$$

If $s \in \bigcup_{i=1}^m (a_i, b_i)$, (i.e. $\phi(s) = 1$) then by (5.2), $P(s) \geq 0$ and since $\theta(s) \leq 1$, we have

$$P(s)\theta(s) \leq P(s)\phi(s).$$

If $s \notin \bigcup_{i=1}^m (a_i, b_i)$, (i.e. $\phi(s) = 0$) then by (5.2), $P(s) \leq 0$ and since $\theta(s) \geq 0$, we also have

$$P(s)\theta(s) \leq P(s)\phi(s).$$

Therefore the last integral in (5.3) is nonpositive which proves

$$\sum_{j=0}^{2n+1} \left(\int_0^1 t^j (\theta(t) - \phi(t))\, dt \right)^2 = 0.$$

This proves Lemma 5.1. □

As a consequence, we deduce

Lemma 5.2 *Let a, b be in $\mathbb{R}$ with $a < b$ and let $\{y_k\}_{k=0}^m$ be a partition of $[a,b]$ of size*

$$\delta = \max_{1\le k\le m} (y_k - y_{k-1}).$$

Let also θ be in $L^\infty(a,b;[0,1])$. Then for every $l \in \mathbb{N}$ there exists $I \subset \{1,\dots,m\}$ such that

$$\tilde{\omega} = \bigcup_{k\in I} (y_{k-1}, y_k), \tag{5.4}$$

satisfies

$$|\tilde{\omega}| \le \int_a^b \theta \, dx, \tag{5.5}$$

$$\left| \int_a^b (\theta - \chi_{\tilde{\omega}})\varphi \, dx \right| \le C(b-a)^{l+1} \|D^{l+1}\varphi\|_{L^1(a,b)} + C\delta \|\varphi\|_{L^\infty(a,b)}, \quad \forall \varphi \in W^{l+1,1}(0,1), \tag{5.6}$$

where C is a positive constant which depends on l but it is independent of θ, δ, a and b.

Proof. It is enough to show the case $a = 0$, $b = 1$. The general one follows using a translation and a dilatation which transforms (a,b) in $(0,1)$.

For a given $l \in \mathbb{N}$, by Lemma 5.1 we know there exists $\omega \subset (0,1)$, satisfying (5.1) and such that the number of discontinuity points of χ_ω in $[0,1]$ is at most $l+1$. We then define

$$I = \{k \in \{1, \cdots, m\} : \ (y_{k-1}, y_k) \subset \omega\}.$$

and $\tilde{\omega}$ by (5.4). By definition of $\tilde{\omega}$, we have $\tilde{\omega} \subset \omega$, and then using (5.1) when $j = 0$ we obtain (5.5). Moreover, using that χ_ω has at most $l+1$ discontinuity points in $[0,1]$, we have

$$|\omega \setminus \tilde{\omega}| \le (l+1)\,\delta. \tag{5.7}$$

We now fix $\varphi \in W^{l+1,1}(0,1)$. Taking a polynomial p of degree l such that

$$\int_0^1 |\varphi - p| \, dx \le C \|D^{l+1}\varphi\|_{L^1(0,1)},$$

with C independent of φ (take for example the Taylor polynomial of degree l of $\varphi \in W^{l+1,1}(0,1) \subset C^l([0,1])$ in some point of $[0,1]$), we get

$$\begin{aligned} \left| \int_0^1 (\theta - \chi_{\tilde{\omega}})\varphi \, dx \right| &\le \left| \int_0^1 (\theta - \chi_\omega)(\varphi - p) \, dx \right| + \left| \int_0^1 (\chi_\omega - \chi_{\tilde{\omega}})\varphi \, dx \right| \\ &\le C \|D^{l+1}\varphi\|_{L^1(0,1)} + (l+1)\delta \, \|\varphi\|_{L^\infty(0,1)}. \end{aligned} \tag{5.8}$$

This proves (5.6) for $a = 0$, $b = 1$. □

Lemma 5.3 *For $r > 0$, small we take a partition $\mathcal{P}^r = \{y_k\}_{k=0}^{m_r}$, with $m_r \in \mathbb{N}$, such that (2.8) is satisfied. We define $\hat{\mathcal{U}}$ by (2.5) and $\mathcal{U}^r$ by (2.10)*

a) For every $\theta \in \hat{\mathcal{U}}$ there exists $\omega \in \mathcal{U}^r$ such that

$$\left| \int_0^x (\theta - \chi_\omega)\varphi \, ds \right| \le C \, r^{\frac{1}{2}} \|\varphi\|_{W^{1,1}(0,1)}, \quad \forall x \in [0,1], \ \forall \varphi \in W^{1,1}(0,1), \tag{5.9}$$

where C is a positive constant independent of θ and r.

b) For every $\theta \in \hat{\mathcal{U}}$ and every $l \in \mathbb{N}$ there exists $\omega \in \mathcal{U}^r$ such that

$$\left| \int_0^1 (\theta - \chi_\omega)\varphi \, ds \right| \le C \, r^{\frac{l+1}{l+2}} \|\varphi\|_{W^{l+1,1}(0,1)}, \quad \forall \varphi \in W^{l+1,1}(0,1), \tag{5.10}$$

where C is a positive constant which depends on l but it is independent of θ and r.

Proof. We take $l \in \mathbb{N}$, $\gamma \in (2r, 1)$ and a subpartition $\mathcal{P}^\gamma = \{z_i\}_{i=0}^{m_\gamma} \subset \mathcal{P}^r$ of $\mathcal{P}^r$ which satisfies

$$\gamma - r \le z_i - z_{i-1} \le \gamma, \ \forall\, i \in \{1, \cdots, m_\gamma - 1\}, \qquad r \le z_{m_\gamma} - z_{m_\gamma - 1} \le \gamma.$$

This implies in particular

$$m_\gamma \le \frac{1}{\gamma - r} + 1 \le \frac{3}{\gamma}. \tag{5.11}$$

Using that for every $i \in \{1, \cdots, m_\gamma - 1\}$ the points y_k with $z_{i-1} \le y_k \le z_i$ are a partition of $[z_{i-1}, z_i]$ with mesh r we can apply Lemma 5.2 in each interval $[z_{i-1}, z_i]$ to construct a set $\omega \in \mathcal{U}$ such that for every $i \in \{1, \cdots, m_\gamma - 1\}$, we have

$$\left| \int_{z_{i-1}}^{z_i} (\theta - \chi_\omega)\varphi \, dx \right| \le C \left(\gamma^{l+1} \|D^{l+1}\varphi\|_{L^1(z_{i-1}, z_i)} + \|\varphi\|_{L^\infty(z_{i-1}, z_i)}\, r \right), \tag{5.12}$$

for every $\varphi \in W^{l+1,1}(0,1)$.

For $x \in [0,1]$, we take the larger j such that $z_j \le x$, then, thanks to (5.12) and (5.11), we have

$$\begin{aligned} \left| \int_0^x (\theta - \chi_\omega)\varphi \, ds \right| &= \left| \int_0^{z_j} (\theta - \chi_\omega)\varphi \, ds + \int_{z_j}^x (\theta - \chi_\omega)\varphi \, ds \right| \\ &\le C\gamma^{l+1} \|D^{l+1}\varphi\|_{L^1(0,1)} + \|\varphi\|_{L^\infty(0,1)} \left(\frac{3r}{\gamma} + (x - z_j) \right). \end{aligned} \tag{5.13}$$

For $l = 0$, the above inequality and $x - z_j < \gamma$ prove

$$\left| \int_0^x (\theta - \chi_\omega)\varphi \, ds \right| \le C\gamma \|D^1\varphi\|_{L^1(0,1)} + C\|\varphi\|_{L^\infty(0,1)} \left(\frac{r}{\gamma} + \gamma \right).$$

Minimizing in γ this quantity we deduce (5.9).

On the other hand, for $x = 1 = z_j$ inequality (5.13) gives

$$\left| \int_0^1 (\theta - \chi_\omega)\varphi \, ds \right| \le C\gamma^{l+1} \|D^{l+1}\varphi\|_{L^1(0,1)} + C\|\varphi\|_{L^\infty(0,1)} \frac{r}{\gamma},$$

which minimizing in γ proves (5.10). □

Using Lemma 5.3 and reasoning similarly to Lemma 4.3, we easily deduce

Lemma 5.4 *Let θ be in $\hat{\mathcal{U}}$ and $f \in L^1(0,1)$. Then, for every $r > 0$ there exists $\omega \in \mathcal{U}^r$ such that, defining u_θ, u_r as the solutions of (2.7) for θ and χ_ω respectively, we have*

a)

$$\|u_\theta - u_r\|_{L^1(0,1)} \le C\big(1 + \|f\|_{L^1(0,1)}\big) r^{\frac{1}{2}}. \tag{5.14}$$

b) If f belongs to $W^{l,1}(0,1)$, then

$$\left\| M_\theta \frac{du_\theta}{dx} - M_{\chi_\omega} \frac{du_r}{dx} \right\|_{L^\infty(0,1)} \le C\big(1 + \|f\|_{W^{l,1}(0,1)}\big) r^{\frac{l+1}{l+2}}. \tag{5.15}$$

Lemma 5.5 *Let $f \in L^1(0,1)$ and θ be in $\hat{\mathcal{U}}$, then for every $r > 0$, there exists $\omega \in \mathcal{U}^r$ such that*

$$|\hat{\mathcal{J}}(\theta) - \mathcal{J}(\omega)| \le C r^{\frac{1}{2}} \big(1 + \|f\|_{L^1(0,1)}\big). \tag{5.16}$$

If for some $l \in \mathbb{N}$ we have that f belongs to $W^{l,1}(0,1)$, $F_1(x,s,\xi)$, $F_2(x,s,\xi)$ are independent of s and belong to $C^{l,1}_{loc}([0,1] \times \mathbb{R})$, then

$$|\hat{\mathcal{J}}(\theta) - \mathcal{J}(\omega)| \le C r^{\frac{l+1}{l+2}} \big(1 + \|f\|_{W^{l,1}(0,1)}\big). \tag{5.17}$$

As a consequence of this Lemma we can now prove

Proof of Theorem 2.7. The existence of solution for problem (1.8) follows from the compactness of $\{\chi_\omega :\ \omega \in \mathcal{U}^r\}$ in $L^1(0,1)$.

The proof of (2.13) is easily deduced from (5.16) with $l = 0$ reasoning as in the proof of Theorem 2.6. Analogously, (2.14) is a consequence of (5.17) and that the functions $F_i(x, s, \xi)$ are supposed independent of s. □

6 An example

In this section we consider a particular case of problem (1.1) for which we can explicitly obtain the optimal control. As a consequence we will give the proof of Proposition 2.8.

Proposition 6.1 *We consider $f \in L^1(0,1)$, f not identically zero, such that*

$$f(t) = f(1-t), \qquad a.e.\ \ t \in [0,1], \tag{6.1}$$

and we define F as the unique primitive function of f satisfying $F(1/2) = 0$.
For $\kappa > 0$, with

$$\kappa \le |\{t \in (0,1) :\ F(t) \neq 0\}| \tag{6.2}$$

and $0 < \alpha < \beta$, we consider the control problem (2.4) corresponding to the functional given by (2.15). Then, the optimal controls for (2.4) are the functions $\theta_0 \in L^\infty(0,1;[0,1])$ which satisfy

$$\int_0^1 \theta_0(t)\,dt = \kappa, \quad \int_0^1 F(t)\theta_0(t)\,dt = 0, \tag{6.3}$$

$$\theta_0(t) = \begin{cases} 1 & \text{if } \ |F(t)| > \gamma_0 \\ 0 & \text{if } \ |F(t)| < \gamma_0, \end{cases} \tag{6.4}$$

with

$$\gamma_0 = \inf\big\{\gamma > 0 :\ \big|\{t \in (0,1) :\ |F(t)| > \gamma\}\big| < \kappa\big\}. \tag{6.5}$$

Proof. Using that for every $\theta \in L^\infty(0,1;[0,1])$ one has

$$\frac{du_\theta}{dt} = \frac{c - F}{M_\theta} \ \text{ in } (0,1),$$

with c defined by

$$\int_0^1 \frac{c-F}{M_\theta} dt = 0 \Longleftrightarrow c = \left(\int_0^1 \frac{1}{M_\theta} dx\right)^{-1} \int_0^1 \frac{F}{M_\theta}\,dx,$$

and that the integral of F in $(0,1)$ vanishes, we have

$$\begin{aligned}
\hat{\partial}(\theta) &= -\int_0^1 M_\theta \left|\frac{du_\theta}{dx}\right|^2 dx = -\int_0^1 \frac{(c-F)(c-F)}{M_\theta}\,dt \\
&= \int_0^1 \frac{F(c-F)}{M_\theta}\,dt = \left(\int_0^1 \frac{dt}{M_\theta}\right)^{-1} \left(\int_0^1 \frac{F}{M_\theta}\,dt\right)^2 - \int_0^1 \frac{|F|^2}{M_\theta}\,dt \\
&= \frac{1}{\alpha\beta}\left((\beta-\alpha)^2 \frac{\left(\int_0^1 F\theta\,dt\right)^2}{\alpha + (\beta-\alpha)\int_0^1 \theta\,dt} - \int_0^1 |F|^2\big(\alpha + (\beta-\alpha)\theta\big)\,dt \right).
\end{aligned}$$

Since the application $(x,y) \in \mathbb{R}\times\mathbb{R}^+ \to x^2/y \in \mathbb{R}$ is convex, we then deduce that $\hat{\mathcal{J}}$ is convex in θ. Moreover, taking into account that F is odd with respect to $1/2$, the above expression shows that given $\theta \in L^\infty(0,1;[0,1])$ and defining $\tilde{\theta} \in L^\infty(0,1;[0,1])$ as

$$\tilde{\theta}(t) = \theta(1-t) \ \text{ a.e. } t \in (0,1),$$

we have

$$\hat{\mathcal{J}}(\theta) = \hat{\mathcal{J}}(\tilde{\theta})$$

and so, by convexity, the symmetrized function θ_0^s of an optimal control θ_0 defined as $\theta_0^s = (\theta_0 + \tilde{\theta}_0)/2$ satisfies

$$\hat{\mathcal{J}}(\theta_0^s) \leq \frac{1}{2}\left(\hat{\mathcal{J}}(\theta_0) + \hat{\mathcal{J}}(\tilde{\theta}_0)\right) = \hat{\mathcal{J}}(\theta_0) \Longrightarrow \hat{\mathcal{J}}(\theta_0^s) = \hat{\mathcal{J}}(\theta_0). \tag{6.6}$$

Using now that for every (x_1,y_1), $(x_2,y_2) \in \mathbb{R}\times\mathbb{R}^+$ one has

$$\frac{\left|\dfrac{x_1+x_2}{2}\right|^2}{\dfrac{y_1}{2}+\dfrac{y_2}{2}} = \frac{1}{2}\frac{|x_1|^2}{y_1} + \frac{1}{2}\frac{|x_2|^2}{y_2} \Longleftrightarrow \frac{x_1}{y_1} = \frac{x_2}{y_2},$$

we deduce that (6.6) implies

$$\frac{\displaystyle\int_0^1 F\theta_0\,dt}{\alpha + (\beta-\alpha)\displaystyle\int_0^1 \theta_0\,dt} = \frac{\displaystyle\int_0^1 F\tilde{\theta}_0\,dt}{\alpha + (\beta-\alpha)\displaystyle\int_0^1 \tilde{\theta}_0\,dt},$$

which using that F is symmetric with respect to $1/2$ is equivalent to

$$\int_0^1 F\theta_0\,dt = 0. \tag{6.7}$$

Therefore, the control problem (2.4) is equivalent to

$$\max_{\theta\in L^\infty(0,1;[0,1])}\left\{\int_0^1 |F|^2\theta\,dt : \ \int_0^1 \theta\,dt \leq \kappa, \ \int_0^1 F\theta\,dt = 0\right\}$$

But thanks to (6.2) is immediate to show that the solutions of problem

$$\max_{\theta\in L^\infty(0,1;[0,1])}\left\{\int_0^1 |F|^2\theta\,dt : \ \int_0^1 \theta\,dt \leq \kappa\right\},$$

are the functions $\theta_0 \in L^\infty(0,1;[0,1])$ which satisfy the first condition in (6.3) and (6.4), and clearly the fact that F is odd with respect to $1/2$ permits to construct functions satisfying these properties and (6.7). This finishes the proof. □

Proof of Proposition 2.8. By Proposition 6.1 problem (2.4) has a unique solution θ_0 given by (this is true for every f which satisfies (6.1), does not changes its sign in $(0,1)$ and it is not the zero function)

$$\theta_0 = \chi_{(0,1/3)\cup(2/3,1)}$$

and

$$\hat{\mathcal{J}}(\theta_0) = -\frac{2}{\alpha}\int_0^{\frac{1}{3}}\left|\frac{1}{2}-t\right|^2 dt - \frac{2}{\beta}\int_{\frac{1}{3}}^{\frac{1}{2}}\left|\frac{1}{2}-t\right|^2 dt. \tag{6.8}$$

Taking

$$k_n = 3\sum_{j=0}^{n-1} 10^j,$$

the same reasoning used in Proposition 6.1 also shows that problem

$$\min_{\theta\in\hat{\mathcal{U}}^n} \hat{\mathcal{J}}$$

has a unique solution θ_0^n given by

$$\theta_0^n(t) = \begin{cases} 1 & \text{if } \ t \in (0, k_n 10^{-n}) \cup (1 - k_n 10^{-n}, 1) \\ \dfrac{1}{3} & \text{if } \ t \in (k_n 10^{-n}, (k_n+1)10^{-n}) \cup (1 - (k_n+1)10^{-n}, 1 - k_n 10^{-n}) \\ 0 & \text{if } \ t \in ((k_n+1)10^{-n}, 1 - (k_n+1)10^{-n}) \end{cases}$$

and

$$\begin{aligned} \hat{\mathcal{J}}(\theta_0^n) = \ & -\frac{2}{\alpha}\int_0^{k_n 10^{-n}} \left|\frac{1}{2} - t\right|^2 dt \\ & -\left(\frac{2}{3\alpha} + \frac{4}{3\beta}\right)\int_{k_n 10^{-n}}^{(k_n+1)10^{-n}} \left|\frac{1}{2} - t\right|^2 dt - \frac{2}{\beta}\int_{(k_n+1)10^{-n}}^{\frac{1}{2}} \left|\frac{1}{2} - t\right|^2 dt. \end{aligned} \tag{6.9}$$

Let us now consider problem

$$\min_{\omega\in\mathcal{U}^n} \mathcal{J}.$$

We have seen in the proof of Proposition 6.1 that the symmetrization θ^s of a function $\theta \in L^\infty(0,1;[0,1])$ satisfies (6.6). This implies that

$$\min_{\omega\in\mathcal{U}^n} \mathcal{J}(\omega) = \min_{\omega\in\mathcal{U}^n} \hat{\mathcal{J}}(\chi_\omega) \geq \min_{\theta\in\mathcal{U}_s^n} \hat{\mathcal{J}}(\theta), \tag{6.10}$$

with

$$\mathcal{U}_s^n = \{\theta \in \hat{\mathcal{U}}^n \ : \ \theta \in \{0, 1/2, 1\}, \text{ a.e. in } (0,1), \ \theta \text{ symmetric with respect to } 1/2\},$$

but using that F is strictly increasing we easily get that the minimum in the right-hand side of (6.10) is attained in a unique function $\theta_0^{n,s}$ defined by

$$\theta_0^{n,s}(t) = \begin{cases} 1 & \text{if } \ t \in (0, k_n 10^{-n}) \cup (1 - k_n 10^{-n}, 1) \\ 0 & \text{if } \ t \in (k_n 10^{-n}, 1 - k_n 10^{-n}). \end{cases}$$

Since this function is a characteristic function, we deduce that the inequality in (6.10) is in fact an equality and

$$\min_{\omega\in\mathcal{U}^n} \mathcal{J}(\omega) = \hat{\mathcal{J}}(\theta_0^n) = -\frac{2}{\alpha}\int_0^{k_n 10^{-n}} \left|\frac{1}{2} - t\right|^2 dt - \frac{2}{\beta}\int_{k_n 10^{-n}}^{\frac{1}{2}} \left|\frac{1}{2} - t\right|^2 dt. \tag{6.11}$$

From (6.8), (6.9) and (6.11) we easily deduce (2.18), (2.19). □

7 Solving the state equation by the finite element method

The purpose of this section is to prove Lemmas 2.11 and 2.15. Lemma 2.11 will permit to estimate the differences (see Corollary 2.13) between control problems (1.10), (1.8) and

the corresponding control problems where the state equations are approximated by the finite element method P^1. Lemma 2.15 provides a counterexample for Lemma 2.11 when the hypothesis $h \leq r$ is removed.

Proof of Lemma 2.11. We know that the solution u_θ of (2.7) is given by (4.11) with c given by (4.13) and g a primitive of f, which we take with zero mean value. Then, we define w as

$$w(x) = \int_0^x \frac{\Pi^h g}{M_\theta} ds - c \int_0^x \frac{1}{M_\theta} ds, \quad \text{a.e. } x \in (0,1), \tag{7.1}$$

with Π^h the operator defined by (4.1) (relative to the partition $\mathcal{P}^h = \{x_k\}_{k=1}^{n_h}$). Then, w is continuous and since θ is constant in each interval (x_{k-1}, x_k) we get that it is affine in each interval (x_{k-1}, x_k). Taking into account that the integral of $\Pi^h g$ coincides with the integral of g in each interval (x_{k-1}, x_k), we get that $w(0) = w(1) = 0$. Therefore, w is in W^h. Moreover, using that in each interval (x_{k-1}, x_k) the integral of $M_\theta \frac{dw}{dx}$ agrees with the one of $M_\theta \frac{dw}{dx}$ we deduce that for every $v \in W^h$ one has

$$\int_0^1 M_\theta \frac{dw}{dx} \frac{dv}{dx}\, dx = \int_0^1 M_\theta \frac{du}{dx} \frac{dv}{dx}\, dx = -\int_0^1 g \frac{dv}{dx}\, dx = \int_0^1 f v\, dx.$$

This proves that w agrees with the solution $\tilde{u}_\theta$ of (2.27).

On the other hand, comparing (4.11) with (7.1) and using that g is in $W^{1,1}(0,1)$ we deduce that

$$\|u_\theta - \tilde{u}_\theta\|_{W^{1,1}(0,1)} \leq C \|f\|_{L^1(0,1)} h.$$

From this inequality u_θ, $\tilde{u}_\theta$ bounded in $W^{1,\infty}(0,1)$ independently of h and the Lipschitz property (2.3) of the functions F_1, F_2 we easily deduce (2.30). □

Proof of Lemma 2.15. Since χ_{ω^k} converges weakly-$*$ in $L^\infty(0,1)$ to θ_0 as k tends to infinity, the first limit in (2.36) is a consequence of Lemma 3.1. Concerning the second limit, note that, in the weak formulation of the discrete problem (2.25), both $\frac{du^h}{dx}$ and $\frac{dv}{dx}$ are constant on each element (x_i, x_{i+1}). Therefore, the left hand side in this weak formulation can be written as

$$\int_0^1 (\alpha \chi_\omega + \beta(1 - \chi_\omega)) \frac{du^h}{dx} \frac{dv}{dx} dx = \int_0^1 \overline{a} \frac{du^h}{dx} \frac{dv}{dx} dx$$

where $\overline{a}$ takes a constant value, in (x_i, x_{i+1}), given by

$$\overline{a}(x) = \frac{1}{h} \int_{x_i}^{x_{i+1}} (\alpha \chi_\omega + \beta(1 - \chi_\omega)) dx, \text{ a.e } x \in (x_i, x_{i+1}).$$

Assume that $h = 1/k$ with $k \in \mathbb{N}$ and let us consider the particular sequence of controls

$$\omega^k = \bigcup_{j=1}^{k} \left(\frac{j-1}{k}, \frac{j - 1/2}{k} \right) \in \mathcal{U}^{h/2}.$$

When considering the particular sequence ω^k, we see that $\overline{a}(x)$ takes the constant value $(\alpha + \beta)/2$ everywhere and for any h. Therefore, the weak formulation in (2.25) coincides with the weak formulation associated to the constant coefficient problem with constant $\bar{a}$ and then, thanks to (2.30),

$$\lim_{k \to \infty} \mathcal{J}^{1/k}(\omega^k) = \lim_{k \to \infty} \hat{\mathcal{J}}^{1/k}(\bar{\theta}) = \hat{\mathcal{J}}(\overline{\theta}),$$

where $\overline{\theta}$ is the constant value such that

$$M(\overline{\theta}) = \frac{\alpha + \beta}{2},$$

i.e. $\overline{\theta} = \alpha/(\alpha + \beta)$ which, in general, is different of $\theta_0 = 1/2$. □

8 Some remarks about the N-dimensional case

Although the aim of the paper is the numerical study of the one-dimensional control problem (1.1), let us give in this section some remarks referred to the N-dimensional problem.

For a bounded open set $\Omega \subset \mathbb{R}^N$, two Carathéodory functions (measurable with respect the first variable and continuous with respect the second and third variables) $F_1, F_2 : \Omega \times \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}$ such that there exist $C > 0$, $h \in L^1(\Omega)$ satisfying

$$|F_1(x,s,\xi)|, |F_2(x,s,\xi)| \leq C\left(h(x) + |s|^2 + |\xi|^2\right), \quad \forall\, (s,\xi) \in \mathbb{R} \times \mathbb{R}^N, \text{ a.e. } x \in \Omega,$$

for a distribution $f \in H^{-1}(\Omega)$ and three positive constants α, β, κ, we consider the control problem

$$\min_{\omega \in \mathcal{U}} \left(\int_\omega F_1(x, u_\omega, \nabla u_\omega)\, dx + \int_{\Omega\setminus\omega} F_2(x, u_\omega, \nabla u_\omega)\, dx \right). \tag{8.1}$$

where, analogously to the control problem (1.1), we have denoted by $\mathcal{U}$ the set

$$\mathcal{U} = \{\omega \subset \Omega : \quad \omega \text{ measurable}, \quad |\omega| \leq \kappa\} \tag{8.2}$$

and by u_ω, for every $\omega \in \mathcal{U}$, the solution of

$$\begin{cases} -\text{div}\, \left(\alpha\chi_\omega + \beta\chi_{\Omega\setminus\omega}\right) \nabla u = f \ \text{ in } \Omega \\ u = 0 \ \text{ on } \partial\Omega. \end{cases} \tag{8.3}$$

As we said in the introduction, problem (8.1) has not a solution in general and so, it is usual to work with a relaxed version of this problem: For $p \in [0,1]$ we denote by $\mathcal{K}(p)$ the set of matrices constructed via homogenization mixing the materials corresponding to the diffusion matrices αI and βI with respective proportions p and $1-p$, and by $\hat{\mathcal{U}}$ (the relaxed control set)

$$\hat{\mathcal{U}} = \left\{(\theta, M) \in L^\infty(\Omega; [0,1]) \times L^\infty(\Omega; \mathbb{R}^{N\times N}) : \quad M \in \mathcal{K}(\theta) \text{ a.e. in } \Omega\right\}. \tag{8.4}$$

It is proved in [5] (see also [2], [3], [4], [9], [16], [18], [21] for related results) that the relaxed control problem is of the form

$$\begin{aligned} &\min \int_\Omega H(x, u, \nabla u, M\nabla u, \theta)\, dx \\ &\begin{cases} -\text{div}\, M\nabla u = f \ \text{ in } \Omega, \qquad u = 0 \ \text{ on } \partial\Omega \\ (\theta, M) \in \hat{\mathcal{U}}, \quad \displaystyle\int_\Omega \theta\, dx \leq \kappa, \end{cases} \end{aligned} \tag{8.5}$$

for a Carathéodory (measurable with respect to the first variable and continuous with respect to the other ones) function H. Some remarks are needed:

Remark 8.1 *As in (2.4), the control θ in (8.5) represents the proportion of material α we are using in the mixture in each point, but now the mixture does not only depend on this proportion but also on the geometric configuration of the materials. Thus, the set $\mathcal{K}(\theta)$ is not reduced to a point as it holds for the one-dimensional problem. In the case we are considering here, corresponding to the optimal mixture of two isotropic materials, an algebraic representation of $\mathcal{K}(\theta)$ is known ([12], [20]). However this does not hold for other interesting problems such as the mixture of more than two materials or the mixture of anisotropic materials. In this sense, it is interesting to remark that in problem (8.5) the matrix M always appears multiplied by ∇u. Thus, problem (8.5) does not permit to*

calculate M but only the product $M\nabla u$. In order to work with (8.5) it is enough to know, for every $\xi \in \mathbb{R}^N$ and $p \in [0,1]$, an explicit characterization of the set

$$\mathcal{K}(p)\xi = \{M\xi \in \mathbb{R}^N : \quad M \in \mathcal{K}(p)\}.$$

In our case, the mixture of two anisotropic materials, $\mathcal{K}(p)\xi$ can be characterized in the following way (this set is known in more general situations, [4], [22]): Denoting by $\lambda(p)$ and $\Lambda(p)$, with $p \in [0,1]$, the harmonic and arithmetic mean of α and β with proportions p and $1-p$, i.e.

$$\lambda(p) = \left(\frac{p}{\alpha} + \frac{1-p}{\beta}\right)^{-1}, \qquad \Lambda(p) = \alpha p + \beta(1-p),$$

we have that $\mathcal{K}(p)\xi$ is the ball

$$\mathcal{K}(p)\xi = \left\{\eta \in \mathbb{R}^N : \quad (\eta - \lambda(p)\xi)\cdot(\eta - \Lambda(p)\xi) \le 0\right\}.$$

Therefore, problem (8.5) can be written in the equivalent form

$$\begin{aligned} &\min \int_\Omega H(x,u,\nabla u,\sigma,\theta)\,dx \\ &\begin{cases} -div\,\sigma = f \quad in\ \Omega, \qquad u = 0 \quad on\ \partial\Omega \\ \theta \in L^\infty(\Omega;[0,1]), \quad \displaystyle\int_\Omega \theta\,dx \le \kappa, \quad (\sigma - \lambda(\theta)\nabla u)\cdot(\sigma - \Lambda(\theta)\nabla u) \le 0 \ \ a.e.\ in\ \Omega. \end{cases} \end{aligned} \tag{8.6}$$

This permits for example to substitute in the definition of the relaxed control set $\hat{\mathfrak{U}}$ the set $\mathcal{K}(p)$ by the (more simple) set of symmetric matrices whose eigenvalues are compressed between $\lambda(p)$ and $\Lambda(p)$.

Remark 8.2 *Defining*

$$E = \left\{(\xi,\eta,p) \in \mathbb{R}^N \times \mathbb{R}^N \times [0,1] : \ (\eta - \lambda(p)\xi)\cdot(\eta - \Lambda(p)\xi) \le 0\right\},$$

the function H which appears in (8.5) is a Carathéodory function with domain $\Omega \times \mathbb{R} \times E$. An explicit expression of H in the whole of its domain is not known in general.

In the particular case where $F_1(x,s,\xi)$, $F_2(x,s,\xi)$ are affine functions in the variable ξ, we have

$$H(x,s,\xi,\eta,p) = pF_1(x,s,\xi) + (1-p)F_2(x,s,\xi), \ \forall\,(s,\xi,\eta,p) \in \mathbb{R}\times E, \quad a.e.\ x \in \Omega,$$

while for nonlinear functions F_i in the variable ξ, an expression of H is only known in some particular cases (which essentially concern with the nonlinear function $|\xi|^2$), see [3], [4], [6], [9] and [16].

However an explicit representation is always known in the boundary of its domain

$$\{(x,s,\xi,\eta,p) :\in \Omega \times \mathbb{R} \times \mathbb{R} \times [0,1] : \ (\eta - \lambda(p)\xi)\cdot(\eta - \Lambda(p)\xi) = 0\},$$

where $H(x,s,\xi,\eta,p)$ is given by

$$\begin{cases} F_1(x,s,\xi) & \text{if } p = 1 \\ F_2(x,s,\xi) & \text{if } p = 0 \\ pF_1\left(x,s,\dfrac{\beta\xi - \eta}{p(\beta-\alpha)}\right) + (1-p)F_2\left(x,s,\dfrac{\eta - \alpha\xi}{(1-p)(\beta-\alpha)}\right) & \text{if } p \ne 0,1. \end{cases} \tag{8.7}$$

Observe that the last line can be taken as the general expression for H, taking the values for $p = 0$ and $p = 1$ by continuity.

Analogously as we did in the one-dimensional case, in order to numerically solve problem (8.5), for $r > 0$ we decompose Ω as

$$\Omega = \bigcup_{i=1}^{m_r} K_i, \quad K_i \text{ disjoint, measurable, } diam(K_i) < r, \quad i \in \{1, \cdots, m_r\}. \tag{8.8}$$

Then, we discretize problem (8.5) as

$$\begin{array}{l} \min \displaystyle\int_\Omega H(x, u, \nabla u, M\nabla u, \theta)\, dx \\ \left\{ \begin{array}{l} -\mathrm{div}\, M\nabla u = f \ \text{ in } \Omega, \qquad u = 0 \ \text{ on } \partial\Omega \\ (\theta, M) \in \hat{\mathfrak{U}}, \quad (\theta, M) \text{ constant in } K_i,\ 1 \le i \le m_r, \quad \displaystyle\int_\Omega \theta\, dx \le \kappa. \end{array} \right. \end{array} \tag{8.9}$$

As we said in Remark 8.2 in the case where the functions $F_i(x, s, \xi)$ are nonlinear in the variable ξ, one of the main difficulties to solve problem (8.9) is that H is not known. To solve this difficulty we can replace H by another function. The following result is proved in [6] in the particular case $F_1(x, s, \xi) = F_2(x, s, \xi) = F(\xi)$. The general case follows similarly:

Theorem 8.3 *We consider a function $\hat{H} : \Omega \times \mathbb{R} \times E \to \mathbb{R} \cup \{+\infty\}$ such that*

$$\hat{H}(., s, \xi, \eta, p) \ \text{ is measurable in } \Omega,\ \forall\, (s, \xi, \eta, p) \in \mathbb{R} \times E \tag{8.10}$$

$$\hat{H}(x, ., ., ., .) \ \text{ is lower semicontinuous in } \mathbb{R} \times E, \ \text{ for a.e. } x \in \Omega \tag{8.11}$$

$$\hat{H}(x, s, \xi, \alpha\xi, 1) = F_1(x, s, \xi), \qquad \hat{H}(x, s, \xi, \beta\xi, 0) = F_2(x, s, \xi) \tag{8.12}$$

$$\hat{H}(x, s, \xi, \eta, p) \ge H(x, s, \xi, \eta, p), \quad \forall (s, \xi, \eta, p) \in \mathbb{R} \times E, \ \text{ a.e. } x \in \Omega. \tag{8.13}$$

For every $r > 0$, we decompose Ω by (8.9). Then, the problem

$$\begin{array}{l} \min \displaystyle\int_\Omega \hat{H}(x, u, \nabla u, M\nabla u, \theta)\, dx \\ \left\{ \begin{array}{l} -div\, M\nabla u = f \ \text{ in } \Omega, \qquad u = 0 \ \text{ on } \partial\Omega \\ (\theta, M) \in \hat{\mathfrak{U}}, \quad (\theta, M) \text{ constant in } K_i,\ 1 \le i \le m_r, \quad \displaystyle\int_\Omega \theta\, dx \le \kappa, \end{array} \right. \end{array} \tag{8.14}$$

has a solution (not unique in general) (θ_r, M_r). Taking u_r as the solution of

$$-div\, M_r \nabla u_r = f \ \text{ in } \Omega, \qquad u_r = 0 \ \text{ on } \partial\Omega,$$

we have

$$\exists \lim_{r \to 0} \int_\Omega \hat{H}(x, u_r, \nabla u_r, M_r \nabla u_r, \theta_r)\, dx = I,$$

with I the minimum value of problem defined by (8.5). The sequence (θ_r, M_r, u_r) is bounded in $L^\infty(\Omega) \times L^\infty(\Omega; \mathbb{R}^{N\times N}) \times H_0^1(\Omega)$. Every function $(\theta, M, u) \in L^\infty(\Omega) \times L^\infty(\Omega; \mathbb{R}^{N\times N}) \times H_0^1(\Omega)$ such that there exists a subsequence of r, still denoted by r, satisfying

$$\theta_r \overset{*}{\rightharpoonup} \theta \ \text{ in } L^\infty(\Omega), \quad M_r \overset{*}{\rightharpoonup} M \ \text{ in } L^\infty(\Omega; \mathbb{R}^{N\times N}), \quad u_r \rightharpoonup u \ \text{ in } H_0^1(\Omega),$$

is such that the function (θ, σ, u), with $\sigma = M\nabla u$ is a solution of (8.6).

Remark 8.4 *A first choice of function $\hat{H}$ is to take*

$$\hat{H}(x,s,\xi,\eta,p)=\begin{cases} F_1(x,s,\xi) & \text{if } p=1,\ \eta=\alpha\xi \\ F_2(x,s,\xi) & \text{if } p=0,\ \eta=\beta\xi \\ +\infty & \text{otherwise.}\end{cases}$$

In this case, taking into account that $\hat{H}(x,u,\nabla u,M\nabla u,\theta)<+\infty$ a.e. in Ω implies that θ is a characteristic function we get that problem (8.14) can be written as

$$\min\left\{\int_\omega F_1(x,u,\nabla u)\,dx+\int_{\Omega\setminus\omega}F_2(x,u,\nabla u)\,dx\right\}$$
$$\begin{cases} -div\,(\alpha\chi_\omega+\beta\chi_\Omega\setminus\omega)\nabla u=f \quad in\ \Omega, \qquad u=0 \quad on\ \partial\Omega \\ \exists\, I\subset\{1,\cdots,m_r\},\ such\ that\ \omega=\displaystyle\bigcup_{i\in I}K_i, \quad |\omega|\le\kappa.\end{cases}$$

Therefore, with this choice of function $\hat{H}$ Theorem 8.3 gives the convergence of the numerical method consisting in discretizing directly the original (unrelaxed) problem (8.1).

Thanks to (8.7), another possibility for $\hat{H}$ is to take $\hat{H}=H$ in $\partial D(H)$, and $\hat{H}=+\infty$, otherwise. For this choice of function $\hat{H}$, taking into account that for $p\neq 0,1$ a matrix $M\in\mathcal{K}(p)$ satisfies

$$\begin{aligned}&(M\xi-\lambda(p)\xi)\cdot(M\xi-\Lambda(p)\xi)=\xi \text{ for some } \xi\neq 0\\ &\Longleftrightarrow M \text{ is a lamination of } \alpha I,\ \beta I \text{ with proportions } p \text{ and } 1-p\\ &\Longleftrightarrow Eig(M)=(\lambda(p),\Lambda(p),\cdots,\Lambda(p)).\end{aligned}$$

we can write problem (8.14) as

$$\min\left\{\int_\Omega\left(pF_1\left(x,u,\frac{\beta\nabla u-M\nabla u}{p(\beta-\alpha)}\right)+(1-p)F_2\left(x,u,\frac{M\nabla u-\alpha\nabla u}{(1-p)(\beta-\alpha)}\right)\right)dx\right\}$$
$$\begin{cases} -div\,M\nabla u=f \quad in\ \Omega, \qquad u=0 \quad on\ \partial\Omega \\ \theta\in L^\infty(\Omega;[0,1]), \quad M \text{ symmetric with } Eig(M)=(\lambda(\theta),\Lambda(\theta),\cdots,\Lambda(\theta)) \text{ a.e. in } \Omega \\ \theta,\ M \text{ constants in } K_i,\ i=1,\cdots,m_r, \quad \displaystyle\int_\Omega\theta\,dx\le\kappa.\end{cases}$$

In this case, problem (8.14) consists in discretizing a partial relaxation of problem (8.1) consisting in considering not only the original controls but also the ones obtained by a simple lamination.

Clearly, when H is known another possibility is to take directly $\hat{H}=H$. In this case we are discretizing the relaxed control problem (8.9).

Remark 8.5 *Although Theorem 8.3 gives the convergence of the discretized problem (8.14), it does not provides any error estimate. In particular, it does not shows what choice of the functions $\hat{H}$ mentioned in Remark 8.4 is better.*

As we saw in the proof of the estimates for the one-dimensional problem, in order to obtain an estimate for the convergence rate of the numerical method, one idea is to construct from a relaxed control (θ,M) another control (θ^r,M^r) in the set of discretized controls such that the solutions of the state equations relative to (θ,M) and (θ^r,M^r) are close. In the case where $\hat{H}=H$ (which can only be used if H is known) one idea is to take (θ^r,M^r) as the mean value of (θ,M) in each element of the triangulation. Denoting by u and u^r the solutions of

$$\begin{cases} -div\,M\nabla u=f \quad in\ \Omega \\ u=0 \quad on\ \partial\Omega,\end{cases}\qquad\qquad \begin{cases} -div\,M^r\nabla u^r=f \quad in\ \Omega \\ u=0 \quad on\ \partial\Omega,\end{cases}$$

with f in $H^{-1}(\Omega)$ and taking into account that

$$-div\, M^r \nabla(u-u^r) = -div\,(M^r - M)\nabla u \quad in\ \Omega,$$

we deduce that

$$\int_\Omega |\nabla(u-u^r)|^2\,dx \le C \int_\Omega |(M^r - M)\nabla u|^2\,dx,$$

which permits to estimate the difference of $u-u^r$ depending of the smoothness properties of M and u and then to estimate the error for the discretized method.

When H is not known and therefore we need to discretize directly the original problem or to consider some partial relaxation the choice of (θ^r, M^r) is not clear.

Remark 8.6 *In Theorem 8.3 we have discretized the set of controls but the state equation is directly solved. It will be interesting to study the convergence when we also discretize this equation and in particular what is the relation we must use between the triangulation chosen for the controls and the one chosen for the resolution of the state equation. A result in this sense can be found in [6], showing that in some cases the method converges using the same triangulation to discretize the controls and the state equation.*

Acknowledgments: The work of the first and third authors was partially supported by the project MTM2008-00306 of the MICINN, Spain and the research group FQM-309 of the CICE, Andalusia.

The work of the second author was partially supported by the Grant MTM2008-03541 of the MICINN, Spain.

The work of the last author was partially supported by the ERC Advanced Grant FP7-246775 NUMERIWAVES, the Grant PI2010-04 of the Basque Government, the ESF Research Networking Programme OPTPDE and Grant MTM2008-03541 of the MICINN, Spain.

The authors are grateful to the Basque Center for Applied Mathematics for its hospitality and support in several visits.

References

[1] G. Allaire. *Shape optimization by the homogenization method.* Appl. Math. Sci. 146, Springer-Verlag, New York, 2002.

[2] G. Allaire, S. Gutiérrez. *Optimal design in small amplitude homogenization.* ESAIM:M2AN 41 (2007), 543-574.

[3] J.C. Bellido, P. Pedregal. *Explicit quasiconvexification for some cost functionals depending on derivatives of the state in optimal designing.* Discr. Contin. Dyn. Syst. 8, 4 (2002), 967-982.

[4] J. Casado-Díaz, J. Couce-Calvo, J.D. Martín-Gómez. *Relaxation of a control problem in the coefficients with a functional of quadratic growth in the gradient.* SIAM J. Control and Optim. 47 (2008), 1428-1459.

[5] J. Casado-Díaz, J. Couce-Calvo, J.D. Martín-Gómez. *Relaxation of optimal design problems with nonlinear functionals in the gradient.* To appear.

[6] J. Casado-Díaz, J. Couce-Calvo, M. Luna-Laynez, J.D. Martín-Gómez. *Optimal design problems for a non-linear cost in the gradient: numerical results.* Applicable Anal. 87 (2008), 1461-1487.

[7] J. Casado-Díaz, J. Couce-Calvo, M. Luna-Laynez, J.D. Martín-Gómez. *Discretization of Coefficient Control Problems with a Nonlinear Cost in the Gradient.* In Integral Methods in Science and Engineering, Volume 2: Computational Methods, ed. by C. Constanda, M.E. Pérez, Birkaüser, Boston, 2010, 55-63.

[8] D. Chenais and E. Zuazua. *Finite Element Approximation of 2D Elliptic Optimal Design.* J. Math. Pures et Appl., 85 (2006), 225-249.

[9] Y. Grabovsky. *Optimal design for two-phase conducting composites with weakly discontinuous objective functionals.* Adv. Appl. Math. 27 (2001), 683-704.

[10] R. Lipton, A.P. Velo. *Optimal design of gradient fields with applications to electrostatics.* In *Nonlinear partial differential equations and their applications*, Collège de France Seminar, Vol. XIV. Eds D. Cioranescu and J. L. Lions, Studies in Math. and its Appl. 31. North-Holland, Amsterdam, 2002, 509-532.

[11] K.A. Lurie. *Applied optimal control theory of distributed systems.* Plenum Press, New York, 1993.

[12] K.A. Lurie, A.V. Cherkaev. *Exact estimates of the conductivity of a binary mixture of isotropic materials.* Proc. Royal Soc. Edinburgh 104 A (1986), 21-38.

[13] F. Murat. *Un contre-example pour le problème du contrôle dans les coefficients.* C.R.A.S Sci. Paris A 273 (1971), 708-711.

[14] F. Murat. *Théorèmes de non existence pour des problèmes de contrôle dans les coefficients.* C.R.A.S Sci. Paris A 274 (1972), 395-398.

[15] F. Murat. *H-convergence.* Séminaire d'Analyse Fonctionnelle et Numérique, 1977-78, Université d'Alger, multicopied, 34 pp. English translation : F. Murat, L. Tartar. *H-convergence.* In *Topics in the Mathematical Modelling of Composite Materials*, ed. by L. Cherkaev, R.V. Kohn. Progress in Nonlinear Diff. Equ. and their Appl., 31, Birkaüser, Boston, 1998, 21-43.

[16] P. Pedregal. *Optimal Design in Two-Dimensional Conductivity for a General Cost Depending on the Field.* Arch. Rat. Mech. Anal., 182, 3 (2006), 367-385.

[17] S. Spagnolo. *Sulla convergenza di soluzioni di equazioni paraboliche ed ellittiche.* Ann. Scuola Norm. Sup. Pisa Cl. Sci., 22, 3 (1968), 571-597.

[18] L. Tartar. *Problèmes de contrôle des coefficients dan des équations aux dérivées partielles.* In *Control theory, numerical methods and computer systems modelling.* Lectures Notes in Econ. and Math. Systems, 10, Springer, Berlin, 1975. Eglish translation : F. Murat, L.Tartar. *On the control of coefficients in partial differential equations.* In *Topics in the Mathematical Modelling of Composite Materials*, ed. by L. Cherkaev, R.V. Kohn. Progress in Nonlinear Diff. Equ. and their Appl., 31, Birkaüser, Boston, 1998, 1-8.

[19] L. Tartar, *Cours Peccot.* Collège de France 1977, partly written in [15].

[20] L. Tartar. *Estimations fines de coefficients homogénéisés.* In *Ennio de Giorgi colloquium (Paris, 1983)*, (ed. P. Kree). Research Notes in Math. 125, Pitman, London, 1985, 168-187.

[21] L. Tartar. *Remarks on optimal design problems.* In *Calculus of variations, homogenization and continuum mechanics,* ed. by G. Buttazzo, G. Bouchitte, P. Suquet. Advances in Math. for Appl Sci, 18, World Scientific, Singapore, 1994, 279-296.

[22] L. Tartar. *An introduction to the homogenization method in optimal design.* In *Optimal shape design*, ed. A. Cellina, A. Ornelas. Lecture Notes in Math. 1740, Springer-Vergag, Berlin, 2000, 47-156.