The Influence of the Tikhonov Term in Optimal Control of Partial Differential Equations

Eduardo Casas

Dedicated to Prof. Enrique Fernández-Cara on the occasion of his 60th birthday.

Abstract In this paper, we analyze the importance of the presence of the Tikhonov term in an optimal control problem. The influence of this term in several aspects of control theory is analyzed: existence and regularity of a solution, convergence of the numerical approximations, and second order optimality conditions.

Keywords Optimal control · Bang-bang controls · State constraints · Second order optimality conditions

1 Introduction

Throughout this paper, Ω denotes an open, bounded subset of $\mathbb{R}^n$, $1 \le n \le 3$, with a Lipschitz boundary Γ, and $0 < T < +\infty$ is fixed. We set $Q = \Omega \times (0, T)$ and $\Sigma = \Gamma \times (0, T)$. Let us consider the following control problem

$$\text{(P)} \quad \min\{J(u) : \alpha \le u(x,t) \le \beta \ \text{ for a.a. } (x,t) \in Q\},$$

where $-\infty \le \alpha < \beta \le +\infty$,

$$J(u) = \frac{1}{2}\int_Q (y_u - y_d)^2\,dx\,dt + \frac{\lambda}{2}\int_Q u^2\,dx\,dt$$

E. Casas (✉)
Departamento de Matemática Aplicada y Ciencias de la Computación, Universidad de Cantabria, Santander, Spain
e-mail: eduardo.casas@unican.es

A. Doubova et al. (eds.), *Recent Advances in PDEs: Analysis, Numerics and Control*, SEMA SIMAI Springer Series 17,
https://doi.org/10.1007/978-3-319-97613-6_5

with $y_d \in L^2(Q)$ and $\lambda \geq 0$. For every control u, we denote y_u the solution of

$$\begin{cases} \dfrac{\partial y}{\partial t} + Ay + a(x,t,y) = u \text{ in } Q, \\ y = 0 \text{ on } \Sigma, \\ y(0) = y_0 \text{ in } \Omega. \end{cases} \tag{1}$$

Here, A is the linear elliptic operator

$$Ay = -\sum_{i,j=1}^{n} \partial_{x_j}[a_{ij}(x)\,\partial_{x_i} y].$$

We make the following assumptions.

Assumption 1 The coefficients $a_{ij} \in L^\infty(\Omega)$ and satisfy

$$\exists \Lambda > 0 \text{ such that } \sum_{i,j=1}^{n} a_{ij}(x)\,\xi_i\,\xi_j \geq \Lambda\,|\xi|^2 \text{ for a.a. } x \in \Omega \text{ and } \forall \xi \in \mathbb{R}^n. \tag{2}$$

Assumption 2 The initial datum $y_0 \in L^\infty(\Omega)$, the target state $y_d \in L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))$, where $\hat{p},\hat{q} \in [2,+\infty]$ are such that $\frac{1}{\hat{p}} + \frac{n}{2\hat{q}} < 1$, and $a : Q \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function of class C^2 with respect to the last variable, satisfying the following assumptions

$$\begin{cases} a(\cdot,\cdot,0) \in L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega)) \text{ and } \exists C_a \leq 0 \text{ such that} \\ \dfrac{\partial a}{\partial y}(x,t,y) \geq C_a \text{ for a.a. } (x,t) \in Q \text{ and } \forall y \in \mathbb{R}, \end{cases} \tag{3}$$

$$\begin{cases} \forall M > 0\ \exists C_M > 0 \text{ such that} \\ \left|\dfrac{\partial^j a}{\partial y^j}(x,t,y)\right| \leq C_M \text{ for a.a. } (x,t) \in Q, \forall |y| \leq M, \text{ with } j = 1,2 \end{cases} \tag{4}$$

$$\begin{cases} \forall \rho > 0 \text{ and } \forall M > 0\ \exists \varepsilon_{M,\rho} > 0 \text{ such that for a.a. } (x,t) \in Q \\ \left|\dfrac{\partial^2 a}{\partial y^2}(x,t,y_2) - \dfrac{\partial^2 a}{\partial y^2}(x,t,y_1)\right| \leq \rho, \forall |y_i| \leq M, \text{ and } |y_2 - y_1| \leq \varepsilon_{M,\rho}. \end{cases} \tag{5}$$

Let us observe that the change of variable $\tilde{y} = e^{C_a t} y$ transforms (1) in

$$\begin{cases} \dfrac{\partial \tilde{y}}{\partial t} + A\tilde{y} + \tilde{a}(x,t,\tilde{y}) = e^{C_a t} u \text{ in } Q, \\ \tilde{y} = 0 \text{ on } \Sigma, \\ \tilde{y}(0) = y_0 \text{ in } \Omega. \end{cases} \tag{6}$$

where $\tilde{a}(x,t,y) = \mathrm{e}^{C_a t} a(x,t,\mathrm{e}^{-C_a t} y) - C_a y$. Now, we infer from (3) that

$$\frac{\partial \tilde{a}}{\partial y}(x,t,y) = \frac{\partial a}{\partial y}(x,t,\mathrm{e}^{C_a t} y) - C_a \geq 0.$$

Then, using the monotonicity of $\tilde{a}$ with respect to y, by classical arguments, we deduce the existence and uniqueness of a solution $\tilde{y}_u \in Y = W(0,T) \cap L^\infty(Q)$ for every $u \in L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))$; see, for instance, [4]. As usual we set

$$W(0,T) = \{y \in L^2(0,T;H_0^1(\Omega)) : \partial_t y \in L^2(0,T;H^{-1}(\Omega))\}.$$

From the equivalence between (1) and (6), we infer the existence and uniqueness of a solution $y_u \in Y$. In fact, we have the following result.

Theorem 1 *Under the above assumptions, for all $u \in L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))$ (1) has a unique solution $y_u \in Y$. The mapping $G : L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega)) \longrightarrow Y$ defined by $G(u) = y_u$ is of class C^2. For all elements u, v, v_1 and v_2 of $L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))$, the functions $z_v = G'(u)v$ and $z_{v_1 v_2} = G''(u)(v_1,v_2)$ are the solutions of the problems*

$$\begin{cases} \dfrac{\partial z}{\partial t} + Az + \dfrac{\partial a}{\partial y}(x,t,y_u)z = v & \text{in } Q, \\ z = 0 & \text{on } \Sigma, \\ z(x,0) = 0 & \text{in } \Omega, \end{cases} \tag{7}$$

and

$$\begin{cases} \dfrac{\partial z}{\partial t} + Az + \dfrac{\partial a}{\partial y}(x,t,y_u)z + \dfrac{\partial^2 a}{\partial y^2}(x,t,y_u) z_{v_1} z_{v_2} = 0 & \text{in } Q, \\ z = 0 & \text{on } \Sigma, \\ z(x,0) = 0 & \text{in } \Omega, \end{cases} \tag{8}$$

respectively.

From this theorem we obtain easily the following result.

Theorem 2 *Under the above assumptions, the functional $J : L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega)) \longrightarrow \mathbb{R}$ is of class C^2. For all u, v, v_1 and v_2 of $L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))$ we have*

$$J'(u)v = \int_Q (\varphi_u + \lambda u) v \, dx \, dt, \tag{9}$$

$$J''(u)(v_1,v_2) = \int_Q \left(1 - \varphi_u \frac{\partial^2 a}{\partial y^2}(x,t,y_u)\right) z_{v_1} z_{v_2} \, dx \, dt, \tag{10}$$

where $z_{v_i} = G'(u)v_i$, $i = 1, 2$, *and* $\varphi_u \in W(0, T) \cap C(\bar{Q})$ *is the solution of*

$$\begin{cases} -\dfrac{\partial \varphi}{\partial t} + A^*\varphi + \dfrac{\partial a}{\partial y}(x, t, y_u)\varphi = y_u - y_d & \text{in } Q, \\ \varphi = 0 & \text{on } \Sigma, \\ \varphi(x, T) = 0 & \text{in } \Omega, \end{cases} \tag{11}$$

For the proof of these theorems, the reader is referred to [8] and [20, Chapter 5]. In the sequel we denote

$$\mathbb{K}_{\alpha,\beta} = \{u \in L^{\hat{p}}(0, T; L^{\hat{q}}(\Omega)) : \alpha \leq u(x, t) \leq \beta \text{ for a.a. } (x, t) \in Q\}.$$

From (9) and the convexity of $\mathbb{K}_{\alpha,\beta}$ we infer the first order optimality conditions satisfied by a local minimum of (P); see, for instance, [20, Section §5.5].

Theorem 3 *Let* $\bar{u} \in \mathbb{K}_{\alpha,\beta}$ *be a local minimum of (P), then there exist elements* $\bar{y} \in Y$ *and* $\bar{\varphi} \in W(0, T) \cap C(\bar{Q})$ *such that*

$$\begin{cases} \dfrac{\partial \bar{y}}{\partial t} + A\bar{y} + a(x, t, \bar{y}) = \bar{u} & \text{in } Q, \\ \bar{y} = 0 & \text{on } \Sigma, \\ \bar{y}(0) = y_0 & \text{in } \Omega, \end{cases} \tag{12}$$

$$\begin{cases} -\dfrac{\partial \bar{\varphi}}{\partial t} + A^*\bar{\varphi} + \dfrac{\partial a}{\partial y}(x, t, \bar{y})\,\bar{\varphi} = \bar{y} - y_d & \text{in } Q, \\ \bar{\varphi} = 0 & \text{on } \Sigma, \\ \bar{\varphi}(T) = 0 & \text{in } \Omega, \end{cases} \tag{13}$$

$$\int_Q (\bar{\varphi} + \lambda\bar{u})(u - \bar{u})\,dx\,dt \geq 0 \quad \forall u \in \mathbb{K}_{\alpha,\beta}. \tag{14}$$

In this paper, we will analyze different issues of the control problem where λ plays a crucial role. The term $\frac{\lambda}{2}\|u\|^2_{L^2(Q)}$ in the cost functional J is called the Tikhonov term, and λ is the Tikhonov parameter. The presence of the Tikhonov term in the cost functional, i.e., $\lambda > 0$, changes very much the control problem: sometimes, it is essential to prove the existence of a solution; it produces a regularizing effect in the optimal control; the second order analysis produces the same results as for finite dimensional optimization problems, with a minimal gap between the necessary and sufficient second order conditions; it is possible to prove good properties of stability of the solutions of (P) with respect to perturbations in the data; we can prove error estimates for the numerical approximation of (P); and, finally, the numerical algorithms work much better when $\lambda > 0$.

When $\lambda = 0$, it is necessary to assume $-\infty < \alpha < \beta < +\infty$ to prove the existence of a solution; the optimal control is essentially discontinuous; the second order analysis is much more complicated and the research is still in progress; in general, it is neither possible to prove stability properties of the solutions of (P) with respect to perturbations of the data, nor can we get error estimates for the numerical approximation; and, finally, the numerical algorithms are unstable.

The first difference in the regularity of the optimal control is deduced from (14). Indeed, if $\lambda > 0$, it is easy to obtain from (14) the identity

$$\bar{u}(x,t) = \mathrm{Proj}_{[\alpha,\beta]}\Big(-\frac{1}{\lambda}\bar{\varphi}(x,t)\Big) \text{ for a.a. } (x,t) \in Q. \tag{15}$$

This implies that $\bar{u}$ inherits some regularity of $\bar{\varphi}$. Actually we have that $\bar{u} \in L^2(0,T;H^1(\Omega)) \cap C(\bar{Q})$. However, for $\lambda = 0$ and $-\infty < \alpha < \beta < +\infty$, (14) leads to

$$\bar{u}(x,t) = \begin{cases} \alpha \text{ if } \bar{\varphi}(x,t) > 0, \\ \beta \text{ if } \bar{\varphi}(x,t) < 0, \end{cases} \quad \text{for a.a. } (x,t) \in Q, \tag{16}$$

which shows that $\bar{u}$ is essentially discontinuous. If the set $\{(x,t) \in Q : \bar{\varphi}(x,t) = 0\}$ has zero Lebesgue measure, then $\bar{u}(x,t) \in \{\alpha,\beta\}$ a.e. in Q. These controls are known in the literature as bang-bang controls. It is frequent for an optimal control to be bang-bang when $\lambda = 0$. More differences will be shown in the rest of the paper.

The plan of this paper is as follows. In Sect. 2, we prove that, unlike it was believed, control constraints are not necessary to establish the existence of a solution of (P) if $\lambda > 0$. Of course, they are necessary if $\lambda = 0$. Moreover, the issue of uniqueness of an optimal control is analyzed. In Sect. 3, we add pointwise state constraints to the control problem. Assuming that $\lambda > 0$, we will prove an extra regularity for the optimal control, improving some existing results. Of course, the assumption $\lambda > 0$ is necessary to prove this additional regularity. In Sect. 4, we analyzed the convergence of the numerical approximation of the control problem. When $\lambda > 0$, the proof of a strong convergence of the controls is well known. However for $\lambda = 0$, only weak convergence is usually established; we prove that the convergence is strong if the continuous control is bang-bang. Finally, in Sect. 5, we show the existing very good results for the second order analysis when $\lambda > 0$, and the difficulties of this analysis when $\lambda = 0$.

2 About the Existence and Uniqueness of Optimal Controls

Theorem 1 establishes the existence of a unique solution y_u for every control $u \in L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))$. If we assume that the controls are only elements of $L^2(Q)$, then the analysis of Eq. (1) is much more involved. Though we could prove the existence of a solution $y_u \in W(0,T)$, this solution does not belong to $L^\infty(Q)$. Hence, the differentiability of the relation $u \in L^2(Q) \to y_u \in W(0,T)$ is not clear at all.

Therefore, the first and second order analysis of the control problem becomes too complicate or even impossible. To overcome this difficulty, it is usual to consider the controls in $L^\infty(Q)$. However, the cost functional J is not coercive in this space, and the existence of a solution to the control problem cannot be proved by standard arguments. This situation is solved by including control constraints of type $\alpha \leq u(x,t) \leq \beta$ with $-\infty < \alpha < \beta < +\infty$ in the formulation of the control problem. In particular, if $\lambda = 0$, the inclusion of control constraints is the only way to ensure the existence of a solution. Here, we prove that it is not necessary to include the control constraints to establish the existence of a solution if $\lambda > 0$. We will also address the issue of uniqueness of solution to (P) in the second part of the section.

2.1 Existence of Solution of (P)

Next, the goal is to prove the existence of a solution of (P). See [12] for a first proof of this result.

Theorem 4 *If* $\lambda > 0$*, then problem (P) has at least one solution* $\bar u$*.*

Proof For every real number $M > 0$ we consider the control problem

$$(\mathrm{P}_M) \quad \min\{J(u) : -M \leq u(x,t) \leq +M \text{ for a.a. } (x,t) \in Q\}.$$

This coincides with (P), where $\alpha = -M$ and $\beta = +M$. The existence of a solution $\bar u_M$ of this problem is proved by taking a minimizing sequence and following the classical arguments. We denote by $\bar y_M$ and $\bar\varphi_M$ the state and adjoint state associated with $\bar u_M$. Then, $(\bar u_M . \bar y_M, \bar\varphi_M)$ satisfies (12)–(14). Now, (15) is written as follows:

$$\bar u_M(x,t) = \mathrm{Proj}_{[-M,+M]}\Big(-\frac{1}{\lambda}\bar\varphi_M(x,t)\Big) \text{ for a.a. } (x,t) \in Q. \tag{17}$$

Since $\bar u_M$ is a solution of (P_M) and 0 is a feasible control of (P_M) we get that $J(\bar u_M) \leq J(0)$, hence

$$\|\bar u_M\|_{L^2(Q)} \leq \frac{1}{\sqrt{\lambda}}\|y^0 - y_d\|_{L^2(Q)} = C_0, \tag{18}$$

where y^0 denotes the state associated with 0. Now, by standard arguments, from (12) we infer with (18)

$$\begin{aligned}\|\bar y_M\|_{L^\infty(0,T;L^2(\Omega))} &\leq C_1\big(\|y_0\|_{L^2(\Omega)} + \|a_0(\cdot,\cdot,0)\|_{L^2(Q)} + \|\bar u_M\|_{L^2(Q)}\big)\\ &\leq C_1\big(\|y_0\|_{L^2(\Omega)} + \|a_0(\cdot,\cdot,0)\|_{L^2(Q)} + C_0\big) = C_2. \end{aligned}\tag{19}$$

Looking at the adjoint state equation (13), we get (see [15, Section §3.7]) with (19)

$$\|\bar{\varphi}_M\|_{L^\infty(Q)} \le C_3\big(\|\bar{y}_M\|_{L^\infty(0,T;L^2(\Omega))} + \|y_d\|_{L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))}\big)$$
$$\le C_3\big(C_2 + \|y_d\|_{L^{\hat{p}}(0,T;L^{\hat{q}}(\Omega))}\big) = C_4. \tag{20}$$

Finally, (17) and (20) imply

$$\|\bar{u}_M\|_{L^\infty(Q)} \le \frac{C_4}{\lambda} = C_\infty \ \forall M > 0. \tag{21}$$

Let us denote by $\bar{u}$ a solution of (P_{M_∞}) for $M_\infty = C_\infty$. We conclude the proof by showing that $\bar{u}$ is a solution of (P). Given an arbitrary element $u \in L^\infty(Q)$ we set $M = \|u\|_{L^\infty(Q)}$. Any solution $\bar{u}_M$ of (P_M) satisfies (21), then it is a feasible control for (P_{M_∞}) and, therefore, $J(\bar{u}) \le J(\bar{u}_M) \le J(u)$. Hence, $\bar{u}$ is a solution of (P). □

2.2 About the Uniqueness of Solution of (P)

Let us observe that the control problem (P) is not convex due to the nonlinearity of the state equation. The uniqueness of a solution is an open question up to now. There is a recent paper [1] where a uniqueness result is proved for a semilinear elliptic control problem under an structural assumption of the nonlinear function a in the state equation, and assuming that λ is large enough. A precise constant η only depending on λ and a is given such that the uniqueness holds whenever the $\|\bar{\varphi}\|_{L^q} \le \eta$ for a certain q depending on a.

If the cost functional is not convex with respect to y, then the uniqueness is false in general. Indeed, let us give an example. We consider the state equation

$$\left\{\begin{aligned} \frac{\partial y}{\partial t} + Ay + y^3 &= u && \text{in } Q,\\ y &= 0 && \text{on } \Sigma,\\ y(0) &= 0 && \text{in } \Omega. \end{aligned}\right. \tag{22}$$

We denote by $z \in Y$ the solution of the problem

$$\left\{\begin{aligned} \frac{\partial z}{\partial t} + Az &= 1 && \text{in } Q,\\ z &= 0 && \text{on } \Sigma,\\ z(0) &= 0 && \text{in } \Omega. \end{aligned}\right. \tag{23}$$

Take λ satisfying

$$0 < \lambda < \frac{1}{|Q|}\int_Q z^2(x,t)\,dx\,dt. \tag{24}$$

Finally, we define the cost functional

$$J(u) = \frac{1}{4}\int_Q (y_u^2 - 1)^2\,dx\,dt + \frac{\lambda}{2}\int_Q u^2\,dx\,dt,$$

and the associated control problem

$$\text{(P)} \quad \min_{u \in L^\infty(Q)} J(u),$$

Since the functional J is not quadratic with respect to y, the existence of a solution of (P) is not a consequence of Theorem 4. However, we can establish the existence of a solution by a small modification of the arguments of the proof of Theorem 4. First, we observe that given $u \in L^\infty(Q)$, (22) has a unique solution $y_u \in L^2(0,T;H_0^1(\Omega)) \cap L^\infty(Q)$. Hence, $\partial_t y_u + Ay_u \in L^2(Q)$ and consequently $y_u \in H^1(Q) \cap C([0,T];H_0^1(\Omega))$; see [19, Section §III.2]. Then, multiplying (22) by y_u^3 we deduce that

$$\|y_u^3\|_{L^2(Q)} \le \|u\|_{L^2(Q)}.$$

Therefore, following again [19] we get

$$\|y_u\|_{L^\infty(0,T;H_0^1(\Omega))} \le C\|u - y_u^3\|_{L^2(Q)} \le 2C\|u\|_{L^2(Q)}.$$

On the other hand, looking at the right hand side of the adjoint state equation

$$\begin{cases} -\dfrac{\partial \varphi_u}{\partial t} + A^*\varphi_u + 3y_u^2\varphi_u = y_u(y_u^2 - 1) & \text{in } Q, \\ \varphi_u = 0 & \text{on } \Sigma, \\ \varphi_u(T) = 0 & \text{in } \Omega, \end{cases} \tag{25}$$

we get

$$\begin{aligned}\|\varphi_u\|_{L^\infty(Q)} &\le C'\|y_u(y_u^2-1)\|_{L^\infty(0,T;L^2(\Omega))} \le C'\big(\|y_u^3\|_{L^\infty(0,T;L^2(\Omega))} + \|y_u\|_{L^\infty(0,T;L^2(\Omega))}\big)\\ &\le C''\big(\|y_u\|^3_{L^\infty(0,T;H_0^1(\Omega))} + \|y_u\|_{L^\infty(0,T;L^2(\Omega))}\big) \le C''\|u\|_{L^2(Q)}\big(8C^3\|u\|^2_{L^2(Q)} + C'''\big).\end{aligned}$$

Using this estimate and arguing as in the proof of Theorem 4 we obtain the existence of a solution $\bar u$ of (P). Let us prove that $\bar u \not\equiv 0$. To this end we observe that (9) and (10) lead to

$$J'(u)v = \int_Q (\varphi_u \lambda u) v \, dx \, dt,$$

$$J''(u)v^2 = \int_Q \left[\left(3y_u^2 - 1 - 6\varphi_u y_u\right) z_v^2 + \lambda v^2\right] dx \, dt,$$

where z_v is the solution of the linearized equation

$$\left\{ \begin{aligned} \frac{\partial z}{\partial t} + Az + 3y_u^2 z = v \quad & \text{in } Q, \\ z = 0 \quad & \text{on } \Sigma, \\ z(0) = 0 \quad & \text{in } \Omega, \end{aligned} \right. \tag{26}$$

For $\bar{u} \equiv 0$ we obtain the state $\bar{y} \equiv 0$. Then, the solution of (25) with $y_u = \bar{y} \equiv 0$ is $\bar{\varphi} \equiv 0$. Hence, the first order necessary optimality condition $J'(\bar{u})v = 0 \ \forall v \in L^\infty(Q)$ holds. However, let us check that the second order necessary condition does not hold. We take $v \equiv 1$. Since $\bar{y} \equiv 0$, we have that the solution z_v of (26) coincides with z, solution of (23). Then, (24) implies

$$J''(\bar{u})v^2 = \int_Q [-z^2(x,t) + \lambda] \, dx \, dt < 0.$$

Hence, the second order necessary condition $J''(\bar{u})v^2 \geq 0 \ \forall v \in L^\infty(Q)$ does not hold. Consequently, $\bar{u} \equiv 0$ is not a solution of (P). Finally, let us take $\tilde{u} = -\bar{u}$. We observe that the associated state $\tilde{y}$ satisfies $\tilde{y} = -\bar{y}$. Therefore, $J(\tilde{u}) = J(\bar{u})$ and $\bar{u} \neq \tilde{u}$, which proves that there exist at least two solutions of (P).

3 Regularity of Optimal Solutions of State-Constrained Control Problems

In this section, we assume that $\lambda > 0$ and include pointwise state constraints.

$$\text{(P)} \quad \min\{J(u) : u \in \mathbb{K}_{\alpha,\beta} \text{ and } a \leq y_u(x,t) \leq b \ \ \forall (x,t) \in \bar{Q}\},$$

where $-\infty < \alpha < \beta < +\infty$ and $-\infty < a < b < +\infty$. Here, we assume that the initial condition y_0 belongs to $C_0(\Omega)$ with

$$C_0(\Omega) = \{z \in C(\bar{\Omega}) : z = 0 \text{ on } \Gamma\}.$$

We also assume that $a < y_0(x) < b \ \forall x \in \bar{\Omega}$.

Due to the continuity of y_0, the fact that $a(\cdot,\cdot,0) \in L^{\hat p}(0,T;L^{\hat q}(\Omega))$ and $u \in L^\infty(Q)$, we have that $y_u \in W(0,T) \cap C(\bar Q)$. This follows from [15, Sections §3.7 and §3.10]. With $M(Q)$ and $M(\Omega)$ we denote the spaces of real and regular Borel measures in Q and Ω, respectively. These space are identified with the dual spaces of $C_0(Q)$ and $C_0(\Omega)$, respectively; see, for instance, [18, Section §6.18]. Analogously to $C_0(\Omega)$, $C_0(Q)$ denotes the space of continuous functions in $\bar Q$ vanishing on ∂Q. Moreover, we have that

$$\|\mu\|_{M(Q)} = \sup\left\{\int_Q z\,d\mu : \|z\|_{C_0(Q)} \le 1\right\} = |\mu|(Q),$$

where $|\mu|(Q)$ is the total variation of μ. The analogous norm is defined in $M(\Omega)$.

Associated with the state constraints we define the set

$$\mathbb{C}_{a,b} = \{z \in C(\bar Q) : a \le z(x,t) \le b\ \forall (x,t) \in \bar Q \text{ and } z = 0 \text{ on } \Sigma\}.$$

Assuming that there exist at least one control $u \in \mathbb{K}_{\alpha,\beta}$ such that the associated state y_u belongs to $\mathbb{C}_{a,b}$, it is easy to prove the existence of a solution of (P). If $\bar u$ is solution of (P), we say that $\bar u$ satisfies the linearized Slater condition if

$$\exists u_0 \in \mathbb{K}_{\alpha,\beta} \text{ such that } a < \bar y(x,t) + z_{u_0-\bar u}(x,t) < b\ \ \forall (x,t) \in \bar Q, \tag{27}$$

where $\bar y = G(\bar u)$ is the state associated with $\bar u$ and $z_{u_0-\bar u} = G'(\bar u)(u_0 - \bar u)$ is the solution of (7) with $y_u = \bar y$ and $v = u_0 - \bar u$. The following optimality conditions are well known [4, 11, 17].

Theorem 5 *If $\bar u$ is a solution of (P) satisfying the linearized Slater condition (27), then there exist $\bar y \in W(0,T) \cap C(\bar Q)$, $\bar\varphi \in L^p(0,T;W_0^{1,q}(\Omega))\ \forall p,q \in [1,2)$ with $\frac{1}{p} + \frac{n}{2q} > \frac{n+1}{2}$, $\bar\mu_Q \in M(Q)$ and $\bar\mu_\Omega \in M(\Omega)$ such that*

$$\begin{cases} \dfrac{\partial \bar y}{\partial t} + A\bar y + a(x,t,\bar y) = \bar u & \text{in } Q,\\ \bar y = 0 & \text{on } \Sigma,\\ \bar y(0) = y_0 & \text{in } \Omega, \end{cases} \tag{28}$$

$$\begin{cases} -\dfrac{\partial \bar\varphi}{\partial t} + A^*\bar\varphi + \dfrac{\partial a}{\partial y}(x,t,\bar y)\,\bar\varphi = \bar y - y_d + \bar\mu_Q & \text{in } Q,\\ \bar\varphi = 0 & \text{on } \Sigma,\\ \bar\varphi(T) = \bar\mu_\Omega & \text{in } \Omega, \end{cases} \tag{29}$$

$$\int_Q (z(x,t) - \bar y(x,t))\,d\bar\mu_Q + \int_\Omega (z(x,T) - \bar y(x,T))\,d\bar\mu_\Omega \le 0\ \ \forall z \in \mathbb{C}_{a,b}, \tag{30}$$

$$\int_Q (\bar\varphi + \lambda\bar u)(u - \bar u)\,dx\,dt \ge 0\ \forall u \in \mathbb{K}_{\alpha,\beta}. \tag{31}$$

We say that $\bar{\varphi} \in L^1(Q)$ is a solution of (29) if

$$\int_Q \bar{\varphi}\Big(\frac{\partial z}{\partial t} + Az + \frac{\partial a}{\partial y}(x,t,\bar{y})z\Big)\,dx\,dt = \int_Q z\,d\bar{\mu}_Q + \int_\Omega z(x,T)\,d\bar{\mu}_\Omega \;\; \forall z \in Z, \tag{32}$$

where

$$Z = \Big\{z \in L^2(0,T;H_0^1(\Omega)) : \frac{\partial z}{\partial t} + Az \in L^\infty(Q) \text{ and } z(x,0) = 0\Big\}.$$

We observe that $Z \subset C(\bar{Q})$ [15, Section §3.7 and §3.10], hence the right hand side of (32) is well defined. In [7], it is proved that there exists a unique solution $\bar{\varphi}$ of (29) in the sense above described, and $\bar{\varphi} \in L^p(0,T;W_0^{1,q}(\Omega))$ $\forall p,q \in [1,2)$ with $\frac{1}{p} + \frac{n}{2q} > \frac{n+1}{2}$. From (31) we deduce again the identity (15). From this identity we infer that $\bar{u} \in L^p(0,T;W^{1,q}(\Omega))$ $\forall p,q \in [1,2)$ with $\frac{1}{p} + \frac{n}{2q} > \frac{n+1}{2}$. For long time, this was the maximal regularity expected for the optimal solution of the control problem. However, by a simple argument that we show below, we obtain that $\bar{u} \in L^2(0,T;H^1(\Omega))$. This additional regularity is very important in the derivation of error estimates for the numerical approximation of the control problem. The proof is based on the following lemma.

Lemma 1 *Let $\bar{\varphi}$ be the solution of (29). Given $M > 0$, we set*

$$\varphi_M(x,t) = \mathrm{Proj}_{[-M,+M]}(\bar{\varphi}(x,t)).$$

Then, $\varphi_M \in L^2(0,T;H_0^1(\Omega))$ and there exists a constant $C = C(\Omega,\Lambda,C_a) > 0$ independent of M such that

$$\|\varphi_M\|_{L^2(0,T;H_0^1(\Omega))} \le C\Big[\|\bar{y} - y_d\|_{L^2(Q)} + \sqrt{M\big(\|\bar{\mu}_Q\|_{M(Q)} + \|\bar{\mu}_\Omega\|_{M(\Omega)}\big)}\Big]. \tag{33}$$

Proof Let us consider two sequences $\{f_k\}_{k=1}^\infty \subset L^2(Q)$ and $\{g_k\}_{k=1}^\infty \subset H_0^1(\Omega)$ satisfying

$$\|f_k\|_{L^1(Q)} \le \|\bar{\mu}_Q\|_{M(Q)} \text{ and } f_k \overset{*}{\rightharpoonup} \bar{\mu}_Q \text{ in } M(Q), \tag{34}$$

$$\|g_k\|_{L^1(\Omega)} \le \|\bar{\mu}_\Omega\|_{M(\Omega)} \text{ and } g_k \overset{*}{\rightharpoonup} \bar{\mu}_\Omega \text{ in } M(\Omega). \tag{35}$$

Now we consider the problem

$$\begin{cases} -\dfrac{\partial \varphi_k}{\partial t} + A^*\varphi_k + \dfrac{\partial a}{\partial y}(x,t,\bar{y})\,\varphi_k = \bar{y} - y_d + f_k & \text{in } Q, \\ \varphi_k = 0 & \text{on } \Sigma, \\ \varphi_k(T) = g_k & \text{in } \Omega. \end{cases} \tag{36}$$

The solution φ_k is unique and belongs to $H^1(Q) \cap C([0,T]; H_0^1(\Omega))$. From (34) and (35) we get the convergence

$$\varphi_k \rightharpoonup \bar\varphi \text{ in } L^p(0,T; W_0^{1,q}(\Omega))\ \forall p,q \in [1,2) \text{ with } \frac{1}{p}+\frac{n}{2q} > \frac{n+1}{2}, \tag{37}$$

$$\lim_{k\to\infty} \|\bar\varphi - \varphi_k\|_{L^r(Q)} = 0\ \ \forall 1 \le r < \frac{n+2}{n}; \tag{38}$$

see [7] for the proof.

Now, we define

$$\varphi_{M,k}(x,t) = \mathrm{Proj}_{[-M,+M]}(\bar\varphi_k(x,t)).$$

Since $\varphi_k \in H^1(Q) \cap C([0,T]; H_0^1(\Omega))$, then $\varphi_{M,k}$ has the same regularity. From the $|\varphi_M(x,t) - \varphi_{M,k}(x,t)| \le |\bar\varphi(x,t) - \varphi_k(x,t)|$ and (38) we infer that $\varphi_{M,k} \to \varphi_M$ strongly in $L^r(Q)$ for every $1 \le r < \frac{n+2}{n}$. If we prove that $\{\varphi_{M,k}\}_{k=1}^\infty$ is bounded in $L^2(0,T; H_0^1(\Omega))$, then the convergence $\varphi_{M,k} \to \varphi_M$ in $L^r(Q)$ implies that $\varphi_M \in L^2(0,T; H_0^1(\Omega))$ as well. To prove this boundedness we multiply Eq. (36) by $\mathrm{e}^{-2C_a t}\varphi_{M.k}$, where C_a was introduced in (3). Then, we get

$$\begin{aligned}&\int_Q -\mathrm{e}^{-2C_a t}\frac{\partial\varphi_k}{\partial t}\varphi_{M,k}\,dx\,dt + \sum_{i,j=1}^n \int_Q \mathrm{e}^{-2C_a t} a_{ij}\,\partial_{x_i}\varphi_k \partial_{x_j}\varphi_{M,k}\,dx\,dt\\ &+ \int_Q \frac{\partial a}{\partial y}(x,t,\bar y)\mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt\\ &= \int_Q \mathrm{e}^{-2C_a t}(\bar y - y_d)\varphi_{M,k}\,dx\,dt + \int_Q \mathrm{e}^{-2C_a t} f_k\varphi_{M,k}\,dx\,dt. \end{aligned}\tag{39}$$

Now using that $\varphi_k\varphi_{M,k} \ge \varphi_{M,k}^2$ and $\varphi_k\partial_t\varphi_{M,k} = \varphi_{M,k}\partial_t\varphi_{M,k} = \frac{1}{2}\partial_t\varphi_{M,k}^2$, we obtain

$$\begin{aligned}&\int_Q -\mathrm{e}^{-2C_a t}\frac{\partial\varphi_k}{\partial t}\varphi_{M,k}\,dx\,dt = -\int_0^T \frac{d}{dt}\int_\Omega \mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt\\ &- 2C_a\int_Q \mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt + \int_Q \mathrm{e}^{-2C_a t}\varphi_k\frac{\partial\varphi_{M,k}}{\partial t}\,dx\,dt\\ &= -\int_\Omega \mathrm{e}^{-2C_a T}\varphi_k(x,T)\varphi_{M,k}(x,T)\,dx + \int_\Omega \varphi_k(x,0)\varphi_{M,k}(x,0)\,dx\\ &- 2C_a\int_Q \mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt + \frac{1}{2}\int_Q \mathrm{e}^{-2C_a t}\partial_t\varphi_{M,k}^2\,dx\,dt. \end{aligned}\tag{40}$$

For the last term we have

$$
\begin{aligned}
&\frac{1}{2}\int_Q \mathrm{e}^{-2C_a t}\partial_t\varphi_{M,k}^2\,dx\,dt\\
&=\frac{1}{2}\int_0^T\frac{d}{dt}\int_\Omega \mathrm{e}^{-2C_a t}\varphi_{M,k}^2\,dx\,dt + C_a\int_0^T\int_\Omega \mathrm{e}^{-2C_a t}\varphi_{M,k}^2\,dx\,dt\\
&\geq \frac{1}{2}\int_\Omega \mathrm{e}^{-2C_a T}\varphi_{M,k}^2(x,T)\,dx - \frac{1}{2}\int_\Omega \varphi_{M,k}^2(x,0)\,dx + C_a\int_Q \mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt\\
&\geq -\frac{1}{2}\int_\Omega \varphi_{M,k}^2(x,0)\,dx + C_a\int_Q \mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt. \qquad (41)
\end{aligned}
$$

From (40) and (41) we deduce

$$
\begin{aligned}
\int_Q -\mathrm{e}^{-2C_a t}\frac{\partial\varphi_k}{\partial t}\varphi_{M,k}\,dx\,dt &\geq -\mathrm{e}^{-2C_a T}\int_\Omega g_k(x)\varphi_{M,k}(x,T)\,dx\\
&+\frac{1}{2}\int_\Omega \varphi_{M,k}^2(x,0)\,dx - C_a\int_Q \mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt\\
&\geq -\mathrm{e}^{-2C_a T}\int_\Omega g_k(x)\varphi_{M,k}(x,T)\,dx - C_a\int_Q \mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt.
\end{aligned}
$$

Inserting this inequality in (39) and using that $\partial_{x_i}\varphi_k\partial_{x_j}\varphi_{M,k} = \partial_{x_i}\varphi_{M,k}\partial_{x_j}\varphi_{M,k}$, we obtain with (2), Young's inequality, (34) and (35)

$$
\begin{aligned}
&\Lambda\int_Q |\nabla\varphi_{M,k}|^2\,dx\,dt + \int_Q\Big[\frac{\partial a}{\partial y}(x,t,\bar{y}) - C_a\Big]\mathrm{e}^{-2C_a t}\varphi_k\varphi_{M,k}\,dx\,dt\\
&\leq \int_Q \mathrm{e}^{-2C_a t}(\bar{y}-y_d)\varphi_{M,k}\,dx\,dt + \int_Q \mathrm{e}^{-2C_a t}f_k\varphi_{M,k}\,dx\,dt + \mathrm{e}^{-2C_a T}\int_\Omega g_k\varphi_{M,k}(T)\,dx\\
&\leq \mathrm{e}^{-2C_a T}\Big[\|\bar{y}-y_d\|_{L^2(Q)}\|\varphi_{M,k}\|_{L^2(Q)} + M\big(\|f_k\|_{L^1(Q)} + \|g_k\|_{L^1(\Omega)}\big)\Big]\\
&\leq C\Big[\|\bar{y}-y_d\|_{L^2(Q)}^2 + M\big(\|\bar{\mu}_Q\|_{M(Q)} + \|\bar{\mu}_\Omega\|_{M(\Omega)}\big)\Big] + \frac{\Lambda}{2}\int_Q |\nabla\varphi_{M,k}|^2\,dx\,dt.
\end{aligned}
$$

Finally, taking into account (3), we get from the above inequality that each $\varphi_{M,k}$ satisfies (33). Hence, φ_M also does it. □

Theorem 6 *Let $\bar{u} \in \mathbb{K}_{\alpha,\beta}$ satisfy (28)–(31). Then, $\bar{u} \in L^2(0,T;H^1(\Omega))$ and the inequality*

$$
\|\bar{u}\|_{L^2(0,T;H^1(\Omega))} \leq C\Big[\|\bar{y}-y_d\|_{L^2(Q)} + \sqrt{M_{\alpha,\beta}\big(\|\bar{\mu}_Q\|_{M(Q)} + \|\bar{\mu}_\Omega\|_{M(\Omega)}\big)}\Big] \qquad (42)
$$

holds, where $C = C(\Omega,\Lambda,C_a,\lambda) > 0$ and $M_{\alpha,\beta} = \max\{|\alpha|,|\beta|\}$.

Proof Let us take $M_{\alpha,\beta}$ as indicated in the statement of the theorem and set

$$\varphi_{M_{\alpha,\beta}}(x,t) = \text{Proj}_{[-M_{\alpha,\beta},+M_{\alpha,\beta}]}(\bar\varphi(x,t)).$$

Then, from Lemma 1 we know that $\varphi_{M_{\alpha,\beta}} \in L^2(0,T;H_0^1(\Omega))$ and

$$\|\varphi_{M_{\alpha,\beta}}\|_{L^2(0,T;H^1(\Omega))} \le C\big[\|\bar y - y_d\|_{L^2(Q)} + \sqrt{M_{\alpha,\beta}\big(\|\bar\mu_Q\|_{M(Q)} + \|\bar\mu_\Omega\|_{M(\Omega)}\big)}\big],$$

for a constant $C = C(\Omega, \Lambda, C_a, \lambda) > 0$. Now, from (31) we have

$$\bar u(x,t) = \text{Proj}_{[\alpha,\beta]}\big(-\frac{1}{\lambda}\bar\varphi(x,t)\big) = \text{Proj}_{[\alpha,\beta]}\big(-\frac{1}{\lambda}\varphi_{M_{\alpha,\beta}}(x,t)\big).$$

This implies that $\bar u \in L^2(0,T;H^1(\Omega))$ as well. Moreover, from the inequality

$$\|\bar u\|_{L^2(0,T;H^1(\Omega))} \le \|\varphi_{M_{\alpha,\beta}}\|_{L^2(0,T;H^1(\Omega))},$$

(42) follows. □

4 Convergence of the Numerical Approximations

In this section, we come back to the problem (P) formulated in Sect. 1 and assume that $-\infty < \alpha < \beta < +\infty$. This problem has at least a solution $\bar u$ for $\lambda \ge 0$. To compute an approximation of $\bar u$ we have to discretize the control problem. The goal in this section is to analyze the convergence of the approximations. The first difficulty of this analysis comes from the convergence of the discretization of the state equation. Here, the difficulty is due to the nonlinear term $a(x,t,y)$ and the low regularity of the solutions y. The main reference for that is [16]. Though the convergence analysis in [16] is carried out for two dimensional domains Ω, Boris Vexler has communicated me a modification of the proof to get a similar result in dimension 3. Using these results and assuming that $\lambda > 0$, then the strong convergence of the controls is proved in a standard way. The idea of the proof is the following. Let $\{u_k\}_{k=1}^\infty$ be a sequence of discrete optimal controls. Since every u_k satisfies the control constraints, the sequence is bound in $L^\infty(Q)$. Hence, we can take a subsequence, denoted in the same way, such that $u_k \overset{*}{\rightharpoonup} \bar u$ in $L^\infty(Q)$, for some control $\bar u$ satisfying the control constraints as well. Now, using [16], we get that the sequence of associated discrete states $\{y_k\}_{k=1}^\infty$ converges strongly to $\bar y$ in $L^2(Q)$, where $\bar y$ is the continuous state associated with $\bar u$. From these convergence properties and using the optimality of every u_k it is easy to prove that $\bar u$ is a solution of (P) and $J(u_k) \to J(\bar u)$. Since $\lambda > 0$, this convergence implies that $\|u_k\|_{L^2(Q)} \to \|\bar u\|_{L^2(Q)}$. This fact and the weak* convergence in $L^\infty(Q)$ imply the strong convergence $u_k \to \bar u$ in $L^2(Q)$. As a consequence of the boundedness of $\{u_k\}_{k=1}^{\infty}$ in $L^\infty(Q)$, we also deduce the strong convergence in $L^p(Q)$ for every $p < +\infty$.

The first part of the above argument can be repeated when $\lambda = 0$ and we obtain that $u_k \overset{*}{\rightharpoonup} \bar{u}$ in $L^\infty(Q)$ and $\bar{u}$ is a solution of (P). Obviously, the above argument to prove the strong convergence fails if $\lambda = 0$. As far as we know, the first result proving the strong convergence of the discrete controls when $\lambda = 0$ was given in [6]. We prove that the convergence of the optimal discrete controls to bang-bang optimal controls is strong. Though this is a very simple exercise, we have confirmed that most of the experts in the field had not realized about this property. The proof is an immediate consequence of the following proposition.

Proposition 1 *Let $\{u_k\}_{k=1}^{\infty}$ be a sequence satisfying $\alpha \le u_k(x,t) \le \beta$ for a.a. $(x,t) \in Q$, and $u_k \overset{*}{\rightharpoonup} \bar{u}$ in $L^\infty(Q)$. Assume that $\bar{u}$ is a bang-bang control. Then, the convergence $u_k \to \bar{u}$ strongly in $L^p(Q)$ holds for every $p < +\infty$.*

Proof Let us denote

$$Q_\alpha = \{(x,t) \in Q : \bar{u}(x,t) = \alpha\} \text{ and } Q_\beta = \{(x,t) \in Q : \bar{u}(x,t) = \beta\}.$$

Since, $\bar{u}$ is a bang-bang control, we have that $|Q| = |Q_\alpha| + |Q_\beta|$, where $|\cdot|$ denotes the Lebesgue measure. Hence, we deduce from the weak* convergence in $L^\infty(Q)$

$$\int_Q |\bar{u} - u_k|\,dx\,dt = \int_{Q_\alpha} (u_k - \bar{u})\,dx\,dt + \int_{Q_\beta} (\bar{u} - u_k)\,dx\,dt$$
$$= \int_Q \chi_{Q_\alpha}(u_k - \bar{u})\,dx\,dt + \int_Q \chi_{Q_\beta}(\bar{u} - u_k)\,dx\,dt \to 0,$$

where χ_{Q_α} and χ_{Q_β} denote the characteristic functions of Q_α and Q_β, respectively. This proves the strong convergence in $L^1(Q)$. Finally, it is enough to observe that

$$\int_Q |\bar{u} - u_k|^p\,dx\,dt \le (\beta - \alpha)^{p-1} \int_Q |\bar{u} - u_k|\,dx\,dt \to 0$$

to conclude the proof. □

5 Second Order Analysis

In this section, we give sufficient second order conditions for local optimality. The main goal is to show the difference between the cases $\lambda > 0$ and $\lambda = 0$. First, let us recall some issues concerning the second order analysis in infinite dimensional spaces. The material presented in this section is based on the papers [8] and [10]; see also [9].

It is well known that second order optimality conditions are an important tool in the numerical analysis of optimization problems. They are essential in proving superlinear or quadratic convergence of numerical algorithms, in deriving error estimates for the numerical discretization of infinite-dimensional optimization problems or just for the proof of local uniqueness of optimal solutions. Although there is an extensive literature on second order optimality conditions, there are still some open problems.

A study of the existing theory of first order optimality conditions reveals that the situation for finite-dimensional problems is very close to the infinite-dimensional one. However, there are big differences when we look at sufficient second order conditions. Let us mention some of these differences.

Consider a differentiable functional $J : U \longrightarrow \mathbb{R}$, where U is a Banach space. If $\bar{u}$ is a local minimum of J, then we know that $J'(\bar{u}) = 0$. This is a necessary condition. If J is not convex, we have to invoke a sufficient condition and should study the second derivative. In the finite-dimensional case, say $U = \mathbb{R}^n$, the first order optimality condition $J'(\bar{u}) = 0$ and the second order condition $J''(\bar{u})v^2 > 0$ for every $v \in U \setminus \{0\}$ imply that $\bar{u}$ is a strict local minimum of J. This second order condition says that the quadratic form $v \to J''(\bar{u})v^2$ is positive definite in $\mathbb{R}^n$, which is equivalent to the strict positivity of the smallest eigenvalue δ_m of the associated symmetric matrix. Moreover, the inequality $J''(\bar{u})v^2 \geq \delta_m \|v\|^2$ for every $v \in \mathbb{R}^n$ holds.

However, if U is an infinite-dimensional space, then the condition $J''(\bar{u})v^2 > 0$ is not equivalent to $J''(\bar{u})v^2 \geq \delta_m \|v\|^2$ for some $\delta_m > 0$. Is one of the two conditions sufficient for local optimality? The next example shows that the first condition is not sufficient for local optimality.

Example 1 Consider the optimization problem

$$(\mathrm{Ex}_1) \quad \min_{u \in L^\infty(0,1)} J(u) = \int_0^1 [tu^2(t) - u^3(t)]\,dt.$$

The function $\bar{u}(t) \equiv 0$ satisfies the first-order necessary condition $J'(\bar{u}) = 0$ and

$$J''(\bar{u})v^2 = \int_0^1 2tv^2(t)\,dt > 0 \ \ \forall v \in L^\infty(0,1) \setminus \{0\}.$$

However, $\bar{u}$ is not a local minimum of (Ex_1). Indeed, if we define

$$u_k(t) = \begin{cases} 2t \text{ if } t \in (0, \frac{1}{k}), \\ 0 \ \text{ otherwise,} \end{cases}$$

then it holds $J(u_k) = -\frac{1}{k^4} < J(\bar{u})$, and $\|u_k - \bar{u}\|_{L^\infty(0,1)} = \frac{2}{k}$.

However, it is well known that if J is of class C^2 in a neighborhood of $\bar{u}$, then the condition $J''(\bar{u})v^2 \geq \delta\|v\|^2\ \forall v \in U$ with $\delta > 0$ is a sufficient condition for local optimality. This seems to solve completely the issue. Nevertheless, this conditions is not so simple in infinite dimensional optimization problems. Let us consider the following example.

Example 2 We discuss the optimization problem

$$(\mathrm{Ex}_2) \quad \min_{u \in L^2(0,1)} J(u) = \int_0^1 \sin(u(t))\, dt.$$

Obviously, $\bar{u}(t) \equiv -\pi/2$ is a global solution. Some fast but formal computations lead to

$$J'(\bar{u})v = \int_0^1 \cos(\bar{u}(t))v(t)\, dt = 0 \ \text{ and}$$

$$J''(\bar{u})v^2 = -\int_0^1 \sin(\bar{u}(t))v^2(t)\, dt = \int_0^1 v^2(t)\, dt = \|v\|^2_{L^2(0,1)}\ \forall v \in L^2(0,1).$$

If the second, stronger condition were sufficient for local optimality, $\bar{u}$ would be strict local minimum of (Ex$_2$). However, this is not true. Indeed, for every $0 < \varepsilon < 1$, the functions

$$u_\varepsilon(t) = \begin{cases} -\dfrac{\pi}{2} & \text{if } t \in [0, 1-\varepsilon], \\[2ex] +\dfrac{3\pi}{2} & \text{if } t \in (1-\varepsilon, 1], \end{cases}$$

are also global solutions of (Ex$_2$), with $J(\bar{u}) = J(u_\varepsilon)$ and $\|\bar{u} - u_\varepsilon\|_{L^2(0,1)} = 2\pi\sqrt{\varepsilon}$. Therefore, infinitely many different global solutions of (Ex$_2$) are contained in any L^2-neighborhood of $\bar{u}$ and $\bar{u}$ is not a strict solution.

What is wrong? The reason is that J is not of class C^2 in $L^2(0,1)$, our fast computations was too careless. Therefore we cannot apply the abstract theorem on sufficient conditions for local optimality in $L^2(0,1)$. On the other hand, J is of class C^2 in $L^\infty(0,1)$ and the derivatives computed above are correct in $L^\infty(0,1)$. However, the inequality $J''(\bar{u})v^2 \geq \delta\|v\|^2_{L^\infty(0,1)}$ does not hold for any $\delta > 0$.

This phenomenon is called the *two-norm discrepancy*: the functional J is twice differentiable with respect to one norm, but the inequality $J''(\bar{u})v^2 \geq \delta\|v\|^2$ holds in a weaker norm in which J is not twice differentiable; see, for instance, [14]. This situation arises frequently in infinite-dimensional problems but it does not happen for finite-dimensions because all the norms are equivalent in this case. The classical theorem on second order optimality conditions can easily be modified to deal with the two norm-discrepancy.

Theorem 7 *Let U be a vector space endowed with two norms, $\| \, \|_\infty$ and $\| \, \|_2$, such that $J : (U, \| \, \|_\infty) \mapsto \mathbb{R}$ is of class C^2 in a neighborhood of $\bar{u}$ and the following properties hold:*

$$J'(\bar{u}) = 0 \quad \textit{and} \ \exists \delta > 0 \ \textit{such that} \ J''(\bar{u})v^2 \geq \delta \|v\|_2^2 \ \forall v \in U, \tag{43}$$

and there exists some $\varepsilon > 0$ such that

$$|J''(\bar{u})v^2 - J''(u)v^2| \leq \frac{\delta}{2}\|v\|_2^2 \ \forall v \in U \ \ \textit{if} \ \ \|u - \bar{u}\|_\infty \leq \varepsilon. \tag{44}$$

Then, there holds

$$\frac{\delta}{4}\|u - \bar{u}\|_2^2 + J(\bar{u}) \leq J(u) \ \ \textit{if} \ \ \|u - \bar{u}\|_\infty \leq \varepsilon \tag{45}$$

so that $\bar{u}$ is a strictly locally optimal with respect to the norm $\| \cdot \|_\infty$.

The proof of this theorem is quite elementary.

Coming back to the control problem (P), we observe that the cost functional, in general, is not of class C^2 in $L^2(Q)$. Hence, the two-norm discrepancy appears in this case. As a consequence, for long time, it was believed that a second order condition of type $J''(\bar{u})v^2 \geq \delta\|v\|^2_{L^2(Q)}$ implied a strict local optimality of $\bar{u}$ in $L^\infty(Q)$. This is a serious drawback in the numerical analysis of (P). More recently, it was proved in [8] that, under the assumption $\lambda > 0$, this condition implies the strict local optimality in $L^2(Q)$ as well; see [2] for a previous, but weaker result, in the case of elliptic control problems. However, the situation is completely different if $\lambda = 0$. Here, we show the difference in the second order analysis between the cases $\lambda > 0$ and $\lambda = 0$. We advance that the second order analysis of (P) with $\lambda > 0$ behaves essentially as the analysis for finite-dimensional optimization problems.

Before a correct formulation of the second order conditions for (P), we need to introduce the cone of critical directions. Given $\bar{u}$ a feasible control satisfying the first order optimality conditions (12)–(14), we define the cone of critical directions

$$C_{\bar{u}} = \left\{ v \in L^2(Q) : v(x,t) \begin{cases} \geq 0 \text{ if } \bar{u}(x,t) = \alpha, \\ \leq 0 \text{ if } \bar{u}(x,t) = \beta, \\ = 0 \text{ if } (\bar{\varphi} + \lambda\bar{u})(x,t) \neq 0, \end{cases} \quad \text{a.e. in } Q \right\}.$$

It is not difficult to prove the necessary second order condition for local optimality $J''(\bar{u})v^2 \geq 0 \ \forall v \in C_{\bar{u}}$; see, for instance, [3, Section §3.2] or [8]. This condition holds independently of the case $\lambda = 0$ or $\lambda > 0$. The differences appear in the statement for the second order sufficient conditions.

5.1 Case λ > 0

We start with the main theorem.

Theorem 8 *Let us assume that $\bar{u}$ is a feasible control for problem (P) satisfying the first order optimality conditions* (12)–(14)*. We also assume that $\lambda > 0$. If the condition $J''(\bar{u})v^2 > 0\ \forall v \in C_{\bar{u}} \setminus \{0\}$ holds, then there exist $\varepsilon > 0$ and $\kappa > 0$ such that*

$$J(\bar{u}) + \frac{\kappa}{2}\|u - \bar{u}\|^2_{L^2(Q)} \le J(u) \quad \forall u \in \mathbb{K}_{\alpha,\beta} \cap \bar{B}_\varepsilon(\bar{u}), \tag{46}$$

where $\bar{B}_\varepsilon(\bar{u})$ denotes the $L^2(Q)$-ball centered at $\bar{u}$ and radius ε.

Proof The reader is referred to, for instance, [8] for a detailed proof. Here, we present a sketch of the proof in order to show the role played by the Tikhonov parameter λ. We argue by contradiction, as in the finite dimensional case. If (46) does not hold, for every integer $k \ge 1$ we deduce the existence of $u_k \in \mathbb{K}_{\alpha,\beta}$ such that

$$\|\bar{u} - u_k\|_{L^2(Q)} < \frac{1}{k} \text{ and } J(\bar{u}) + \frac{1}{2k}\|\bar{u} - u_k\|^2_{L^2(Q)} > J(u_k) \quad \forall k \ge 1. \tag{47}$$

We set $\rho_k = \|\bar{u} - u_k\|_{L^2(Q)}$ and $v_k = \frac{1}{\rho_k}\|\bar{u} - u_k\|_{L^2(Q)}$. By taking a subsequence if necessary, we can assume that $v_k \rightharpoonup v$ in $L^2(Q)$. The proof is split into three steps.

Step 1: $v \in C_{\bar{u}}$ It is obvious that the set

$$S = \left\{ v \in L^2(Q) : v(x,t) \begin{cases} \ge 0 \text{ if } \bar{u}(x,t) = \alpha, \\ \le 0 \text{ if } \bar{u}(x,t) = \beta, \end{cases} \text{a.e. in } Q \right\}$$

is convex and closed in $L^2(Q)$. Moreover, since $\alpha \le u_k(x,t) \le \beta$ a.e. in Q $\forall k \ge 1$, we deduce that $\{v_k\}_{k=1}^\infty \subset S$. Hence, $v \in S$ holds. It remains to prove that $v(x,t) = 0$ if $(\bar{\varphi} + \lambda\bar{u})(x,t) \ne 0$ a.e. in Q. First, we observe that (14) implies that

$$\bar{u}(x,t) = \begin{cases} \alpha \text{ if } (\bar{\varphi} + \lambda\bar{u})(x,t) > 0, \\ \beta \text{ if } (\bar{\varphi} + \lambda\bar{u})(x,t) < 0, \end{cases} \quad \text{for a.a.} (x,t) \in Q. \tag{48}$$

This implies that $(\bar{\varphi} + \lambda\bar{u})(x,t)w(x,t) \ge 0$ a.e. in Q $\forall w \in S$. Therefore it also holds for $w = v$. Now, from (47) we infer

$$\frac{J(\bar{u} + \rho_k v_k) - J(\bar{u})}{\rho_k} = \frac{J(u_k) - J(\bar{u})}{\rho_k} < \frac{1}{2k}\|u_k - \bar{u}\|_{L^2(Q)}.$$

Passing to the limit when $k \to \infty$ we get

$$\int_Q |\bar\varphi + \lambda\bar u||v|\,dx\,dt = \int_Q (\bar\varphi + \lambda\bar u)v\,dx\,dt = \lim_{k\to\infty} \frac{J(\bar u + \rho_k v_k) - J(\bar u)}{\rho_k} \le 0.$$

This concludes the proof of $v \in C_{\bar u}$.

Step 2: $J''(\bar u)v^2 \le 0$ Using again (47), the fact that $J'(\bar u)w \ge 0\ \forall w \in C_{\bar u} \subset S$, and making a Taylor expansion of $J(u_k)$ around $\bar u$, we get

$$\begin{aligned}\frac{\rho_k^2}{2k} &= \frac{1}{2k}\|u_k - \bar u\|^2_{L^2(Q)} > J(u_k) - J(\bar u) = J(\bar u + \rho_k v_k) - J(\bar u)\\ &= \rho_k J'(\bar u)v_k + \frac{\rho_k^2}{2}J''(\bar u + \theta_k\rho_k v_k)v_k^2 \ge \frac{\rho_k^2}{2}J''(\bar u + \theta_k(u_k - \bar u))v_k^2.\end{aligned}$$

Dividing the above inequality by $\frac{\rho_k^2}{2}$ we obtain: $J''(\bar u+\theta_k(u_k-\bar u))v_k^2 < 1/k$. Hence, passing to the limit when $k \to \infty$, we conclude that $J''(\bar u)v^2 \le 0$.

Step 3: Contradiction The second order condition $J''(\bar u)v^2 > 0\ \forall v \in C_{\bar u} \setminus \{0\}$ along with the proved steps 1 and 2 implies that $v_k \rightharpoonup v = 0$. Let us set $\hat u_k = \bar u + \theta_k(u_k - \bar u)$, $\hat y_k$ its associated state, $\hat\varphi_k$ the corresponding adjoint state, and $\hat z_k$ the solution of (7), where u is replaced by $\hat u_k$ and v by v_k. The weak convergence $v_k \rightharpoonup v$ in $L^2(Q)$ implies the strong convergence $\hat z_k \to 0$ in $L^2(Q)$. Then, we deduce from (10) and the fact that $\|v_k\|_{L^2(Q)} = 1\ \forall k \ge 1$

$$0 < \lambda = \lim_{k\to\infty}\int_Q \Big[\big(1 - \hat\varphi_k\frac{\partial^2 a}{\partial y^2}(x,t,\hat y_k)\big)\hat z_k^2 + \lambda v_k^2\Big]\,dx\,dt = \lim_{k\to\infty} J''(\hat u_k)v_k^2 = J''(\bar u)v^2 \le 0,$$

which is a contradiction. □

Observe that it was not required a second order condition of type $J''(\bar u)v^2 \ge \delta\|v\|^2_{L^2(Q)}\ \forall v \in C_{\bar u}$ as one can expect for an infinite dimensional optimization problem. The issue is that this second order condition is equivalent to $J''(\bar u)v^2 > 0\ \forall v \in C_{\bar u}$. Once again, this equivalence is valid just because $\lambda > 0$. In fact the following result holds (see [8, 9]).

Theorem 9 *Let us assume that* $\lambda > 0$. *Then, the following statements are equivalent*

$$J''(\bar u)v^2 > 0\ \forall v \in C_{\bar u} \setminus \{0\}, \tag{49}$$

$$\exists\delta > 0 \text{ and } \tau > 0 \text{ such that } J''(\bar u)v^2 \ge \delta\,\|v\|^2_{L^2(Q)}\ \forall v \in C^\tau_{\bar u}, \tag{50}$$

$$\exists\delta > 0 \text{ and } \tau > 0 \text{ such that } J''(\bar u)v^2 \ge \delta\|z_v\|^2_{L^2(Q)}\ \forall v \in C^\tau_{\bar u}, \tag{51}$$

where z_v is the solution of (7) *corresponding to $y_u = \bar{y}$ and*

$$C_{\bar{u}}^{\tau} = \left\{ v \in L^2(Q) : v(x,t) \begin{cases} \geq 0 \text{ if } \bar{u}(x,t) = \alpha, \\ \leq 0 \text{ if } \bar{u}(x,t) = \beta, \\ = 0 \text{ if } |(\bar{\varphi} + \lambda\bar{u})(x,t)| \geq \tau, \end{cases} \quad \text{a.e. in } Q \right\}.$$

Observe that $C_{\bar{u}}$ is strictly contained in $C_{\bar{u}}^{\tau}$ $\forall\tau > 0$. Therefore (50) seems to be a stronger condition that the usual one: $J''(\bar{u}) \geq \delta\|v\|_{L^2(Q)}^2$ $\forall v \in C_{\bar{u}}$. But, actually they are equivalent as follows from the previous theorem.

5.2 Case $\lambda = 0$

When $\lambda = 0$, the proof of Theorem 8 fails precisely at the last step. There is no way to get the contradiction. In this case, Theorem 9 is also false. Since, the condition (49) is not enough to prove that $\bar{u}$ is a local minimum, one can try to check if the condition (50) is sufficient for a local optimality. However, looking at the expression of $J''(\bar{u})v^2$

$$J''(u)v^2 = \int_Q \left[\left(1 - \bar{\varphi}\frac{\partial^2 a}{\partial y^2}(x,t,\bar{y})\right)z_v^2\right] dx\, dt,$$

this condition seems quite difficult to be fulfilled. In fact, it was proved in [5] that it does not hold, except maybe in a few extreme cases. Finally, the condition (51) makes sense if we compare with the second derivative $J''(\bar{u})v^2$. In [5], it was proved that (51) is a sufficient second order optimality condition. More precisely, assuming that $\bar{u} \in \mathbb{K}_{\alpha,\beta}$ satisfies (12)–(14) and (51), then there exist $\varepsilon > 0$ and $\kappa > 0$ such that

$$J(\bar{u}) + \frac{\kappa}{2}\|y_u - \bar{y}\|_{L^2(Q)}^2 \leq J(u) \quad \forall u \in \mathbb{K}_{\alpha,\beta} \cap B_\varepsilon(\bar{u}), \tag{52}$$

where $B_\varepsilon(\bar{u})$ denotes again the $L^2(Q)$-ball.

Unlike (46), the above inequality does not allow to prove, in general, neither stability of the optimal control with respect to perturbations of the data of the control problem, nor we can derive error estimates in the control for the numerical approximation. However, they are useful to prove stability of the states or to get error estimates for the states.

The main drawback of the condition (51) is that the gap with the necessary second order optimality condition is big. However, for $\lambda > 0$, this gap is minimal, in fact, the same as in finite dimension. Recently, some results have been obtained for bang-bang controls where the gap is smaller; see [13]. However, the problem is not completely solved and some research is in progress.

Acknowledgements The author was partially supported by Ministerio Español de Economía, Industria y Competitividad under the research projects MTM2014-57531-P and MTM2017-83185-P.

References

1. Ali, A.A., Deckelnick, K., Hinze, M.: Global minima for semilinear optimal control problems. Comput. Optim. Appl. **65**, 261–288 (2016)
2. Bonnans, J.F.: Second-order analysis for control constrained optimal control problems of semilinear elliptic systems. Appl. Math. Optim. **38**, 303–325 (1998)
3. Bonnans, F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, Berlin (2000)
4. Casas, E.: Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations. SIAM J. Control Optim. **35**(4), 1297–1327 (1997)
5. Casas, E.: Second order analysis for bang-bang control problems of PDEs. SIAM J. Control Optim. **50**(4), 2355–2372 (2012)
6. Casas, E., Chrysafinos, K.: Error estimates for the approximation of the velocity tracking problem with bang-bang controls. ESAIM Control Optim. Calc. Var. **23**, 1267–1291 (2017)
7. Casas, E., Kunisch, K.: Parabolic control problems in space-time measure spaces. ESAIM Control Optim. Calc. Var. **22**(2), 355–370 (2016)
8. Casas, E., Tröltzsch, F.: Second order analysis for optimal control problems: improving results expected from abstract theory. SIAM J. Optim. **22**(1), 261–279 (2012)
9. Casas, E., Tröltzsch, F.: Second order optimality conditions and their role in PDE control. Jahresber. Dtsch. Math. Ver. **117**(1), 3–44 (2015)
10. Casas, E., Tröltzsch, F.: Second order optimality conditions for weak and strong local solutions of parabolic optimal control problems. Vietnam J. Math. **44**(1), 181–202 (2016)
11. Casas, E., Raymond, J.P., Zidani, H.: Pontryagin's principle for local solutions of control problems with mixed control-state constraints. SIAM J. Control Optim. **39**(4), 1182–1203 (2000)
12. Casas, E., Mateos, M., Rösch, A.: Finite element approximation of sparse parabolic control problems. Math. Control Relat. Fields **7**(3), 393–417 (2017)
13. Casas, E., Wachsmuth, D., Wachsmuth, G.: Sufficient second-order conditions for bang-bang control problems. SIAM J. Control Optim. **55**(5), 3066–3090 (2017)
14. Ioffe, A.D.: Necessary and sufficient conditions for a local minimum. III. Second order conditions and augmented duality. SIAM J. Control Optim. **17**(2), 266–288 (1979). MR 525027 (82j:49005c)
15. Ladyzhenskaya, O.A., Solonnikov, V.A., Ural'tseva, N.N.: Linear and Quasilinear Equations of Parabolic Type. American Mathematical Society, Providence (1988)
16. Neitzel, I., Vexler, B.: A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems. Numer. Math. **120**, 345–386 (2012)
17. Raymond, J.P., Zidani, H.: Pontryagin's principle for state-constrained control problems governed by parabolic equations with unbounded controls. SIAM J. Control Optim. **36**(6), 1853–1879 (1998)
18. Rudin, W.: Real and Complex Analysis. McGraw-Hill, London (1970)
19. Showalter, R.E.: Monotone Operators in Banach Space and Nonlinear Partial Differential Equations. Mathematical Surveys and Monographs, vol. 49. American Mathematical Society, Providence (1997). MR 1422252 (98c:47076)
20. Tröltzsch, F.: Optimal Control of Partial Differential Equations: Theory, Methods and Applications. Graduate Studies in Mathematics, vol. 112. American Mathematical Society, Philadelphia (2010)