Dissertation

Metric Regularity in Model Predictive and Optimal Control

carried out for the purpose of obtaining the degree of Doctor technicae (Dr. techn.), submitted at TU Wien, Faculty of Mathematics and Geoinformation, by

Alberto Domínguez Corella MSc.
Mat.Nr.: 11936035

under supervision of
Prof. Dr. Vladimir Veliov
Institute of Statistics and Mathematical Methods in Economics

Vienna, November 2022

REVIEWERS

Prof. Dr. Fredi Tröltzsch
Technische Universität Berlin
Institut für Mathematik
Sekretariat MA 4-4 Straße des 17. Juni 136, Berlin, Germany

Prof. Dr. Arnd Rösch
Universität Duisburg-Essen
Fakultät für Mathematik
Thea-Leymann-Straße 9 D-45127 Essen, Germany

Affidavit

I declare in lieu of oath, that I wrote this thesis and performed the associated research myself, using only literature cited in this volume. If text passages from sources are used literally, they are marked as such.

I confirm that this work is original and has not been submitted elsewhere for any examination, nor is it currently under consideration for a thesis elsewhere.

I acknowledge that the submitted work will be checked electronically-technically using suitable and state-of-the-art means (plagiarism detection software). On the one hand, this ensures that the submitted work was prepared according to the high-quality standards within the applicable rules to ensure good scientific practice "Code of Conduct" at the TU Wien. On the other hand, a comparison with other student theses avoids violations of my personal copyright.

Alberto Domínguez Corella
Vienna, November 2022

Abstract

The stability of optimal solutions is a central issue in control theory. In this thesis we study sufficient conditions for various notions of stability in the context of optimal control. Under some structural assumptions, an optimal control problem can be associated with a set-valued mapping, the so-called optimality mapping, which represents the first-order necessary conditions of the problem. The concept of metric regularity abstracts the different notions of stability into properties of the optimality mapping. This thesis studies enhanced versions of the known metric regularity properties and their interaction with optimal control problems; we also illustrate the role of stability with some applications in numerical analysis. This work is cumulative; it consists of an introduction and four published or accepted journal articles. The introduction is intended as a complement to the papers, providing adequate preliminaries as well as a summary of the results obtained in the papers.

The first paper studies a stronger version of the so-called strong metric regularity property in optimal control of ordinary differential equations. This notion allows to study stronger notions of stability than the previous ones in the literature, and is suitable for a class of problems (affine with respect to the control variable) for which the standard assumptions do not hold. An example is presented as well as an application to numerical analysis, namely the so-called uniform Euler discretization method. The second paper goes further in the error analysis of numerical methods for nonlinear optimal control problems and studies the accuracy, by means of the metric subregularity property, of the so-called Model Predictive Control (MPC) algorithm, a well-established and widely approach for generating feedback control strategy. Finally, the third paper and the fourth one are both dedicated to the stability analysis of optimal control problems constrained by elliptic partial differential equations. The third one studies in a general framework the subregularity property, focusing on sufficient conditions and applications to the stability of optimal solutions. In the fourth one, new assumptions are introduced in the literature in form of second order sufficient conditions for optimality; moreover, it is proved that this assumptions are enough for stability of the optimal states. Though the subregularity property is not explicitly mentioned in the last paper, all of the methods used there come from the understanding of it and its corresponding translation to the stability analysis of the optimal states.

Kurzfassung

Die Stabilität von optimalen Lösungen ist ein zentrales Thema in der Kontrolltheorie. In dieser Thesis untersuchen wir hinreichende Bedingungen für verschiedene Stabilitätsbegriffe im Kontext optimaler Kontrolletheorie. Unter einigen strukturellen Annahmen kann ein Optimalsteuerungsproblem mit einer mengenwertigen Abbildung, der sogenannten Optimalitätsabbildung, assoziiert werden, die die notwendigen Bedingungen erster Ordnung des Problems darstellt. Das Konzept der metrischen Regularität abstrahiert die verschiedenen Stabilitätsbegriffe in Eigenschaften der Optimalitätsabbildung. Diese Dissertation untersucht erweiterte Versionen der bekannten metrischen Regularitätseigenschaften und deren Wechselwirkung mit Optimalsteuerungsproblemen; Wir veranschaulichen auch die Rolle der Stabilität mit einigen Anwendungen in der numerischen Analyse. Diese Arbeit ist kumulativ; Es besteht aus einer Einleitung und vier Artikeln. Die Einführung ist als Ergänzung zu den Artikeln gedacht, indem sie sowohl eine angemessene Einleitung als auch eine Zusammenfassung der in den Artikeln erzielten Ergebnisse liefert.

Die erste Arbeit [14] untersucht eine stärkere Version der sogenannten *strong metric regularity* bei der optimalen Kontrolle von gewöhnlichen Differentialgleichungen. Dieser Begriff ermöglicht es, stärkere Stabilitätsbegriffe als die vorherigen in der Literatur zu untersuchen, und eignet sich für eine Klasse von Problemen (affine), für die die Standardannahmen nicht gelten. Ein Beispiel wird vorgestellt sowie eine Anwendung auf die numerische Analysis, das sogenannte *uniform Euler discretization scheme*. Die zweite Arbeit [1] geht weiter in die Fehleranalyse numerischer Verfahren für nichtlineare Optimalsteuerungsprobleme und untersucht die Genauigkeit mittels der metrischen Subregularitätseigenschaft (eine erweiterte Version) der sogenannten Model Predictive Control (MPC)-Algorithmus, eine gut etablierte und weit verbreitete Feedback-Control-Strategie. Die dritte [13] und die vierte Arbeit [12] schließlich widmen sich beide der Stabilitätsanalyse von Optimalsteuerungsproblemen, die durch elliptische partielle Differentialgleichungen eingeschränkt sind. Die dritte untersucht in einem allgemeinen Rahmen die Subregularitätseigenschaft, wobei der Schwerpunkt auf ihren hinreichenden Bedingungen und ihren Anwendungen auf die Stabilität von Kontrolllösungen liegt. In der vierten werden neue Annahmen in Form von hinreichenden Bedingungen zweiter Ordnung für Optimalität in die Literatur eingeführt; außerdem ist bewiesen, dass diese Annahmen für die Stabilität optimaler Zustände hinreichend sind. Obwohl die Subregularitätseigenschaft im letzten Artikel nicht explizit erwähnt wird, stammen alle dort verwendeten Methoden aus ihrem Verständnis und ihrer entsprechenden Übersetzung in die Stabilitätsanalyse der optimalen Zustände.

Contents

Chapter 1

Introduction

One of the central issues in optimization is the error analysis of methods, algorithms and techniques applied to the calculation of optimal solutions. To get a quick picture of what this work is about, suppose that we want to solve a given problem, but it is hard to do so. We might resort to solve "approximate" problems, simplified version perhaps. If the approximations to the problem are consistent with the structure of the original problem, one can expect that solutions of the approximate problems "converge" (in some sense) to the solution of the original problem. Of course, difficulties may arise, such as that the solutions may not be unique, or they may not even exist, etc. It is here when the notion of stability comes to mind, it is then clear that this should be translated into a mathematical framework. The notion of continuity can, in many cases, deal with the qualitative aspects of problems re-cast in terms of functions. However, in many applications, a quantitative analysis is needed to understand the convergence relation of the approximate problems and the required computational efforts. This quantitative analysis of stability can be abstracted into properties of (set-valued) functions; the so-called metric regularity properties. Under some assumptions, an optimal control problem can be associated with a set-valued mapping, the so-called optimality mapping, which represents the first-order necessary conditions of the problem. Instead of investigating regularity properties of an optimal control problem, we do this for the optimality mapping, which takes the form of a (generalized) equation.

Metric regularity theory lies in the very heart of variational analysis, it can be consider a relatively new discipline since it started in the 80's. It arose largely out of the needs of modern optimization theory, in which phenomena such as non-differentiability and set-valued mappings naturally appear. As we will see along the introduction, the concept of metric regularity connects all the four papers in this thesis.

The chronological order of the papers is the following.

- Domínguez Corella, Alberto and Quincampoix, Marc and Veliov, Vladimir M.: Strong bi-metric regularity in affine optimal control problems. Pure Appl. Funct. Anal., 6(6):1119-1137,2021
- Angelov, Georgi and Domínguez Corella, Alberto and Veliov, Vladimir M.: On the accuracy of the model predictive control method. SIAM Journal on Control and Optimization, 60(4):2469-2487, 2022.
- Domínguez Corella, Alberto and Jork, Nicolai and Veliov, Vladimir M. Stability in affine optimal control problems constrained by semilinear elliptic partial differential equations. To appear in ESAIM. Control, Optimisation and Calculus of Variations.
- Casas, Eduardo and Domínguez Corella, Alberto and Jork, Nicolai. New assumptions for stability analysis in elliptic optimal control problems. To appear in SIAM Journal of Control and Optimization.

The first paper [14] studies a stronger version of the so-called strong metric regularity property in optimal control of ordinary differential equations. This notion allows to study stronger notions of stability than the previous ones in the literature, and is suitable for a class of problems (affine ones) for which the standard assumptions do not hold. An example is presented as well as an application to numerical analysis, the so-called uniform Euler discretization method. The second paper [1] goes further in the error analysis of numerical methods for nonlinear optimal control problems and studies the accuracy, by means of the metric subregularity property (an enhanced version), of the so-called Model Predictive Control (MPC) algorithm, a well-established and widely used feedback control strategy. Finally, the third paper [13] and the fourth one [12] are both dedicated to the stability analysis of optimal control problems constrained by elliptic partial differential equations. The third one studies in a general framework the subregularity property, focusing in its sufficient conditions and its applications to the stability of control solutions. In the fourth one, new assumptions are introduced in the literature in form of second order sufficient conditions for optimality; moreover, it is proved that this assumptions are enough for stability of optimal states. Though the subregularity property is not explicitly mentioned in the paper, all of the methods used there come from the understanding of it and its corresponding translation to the stability analysis of the optimal states. Another important feature of the problem studied in [12] is that we consider a nonmonote non-coercive elliptic equation of the type studied in [4, 5].

This introduction intends to be complementary to the main text (the papers) as well a brief summary of the main results. We start motivating and reviewing some of the metric regularity properties in the literature. After that, we will introduce the so-called Lagrange problem in nonlinear optimal control of ordinary differential equations. With the adequate optimal control preliminaries at hand, we will proceed to introduce the Model Predictive Control (MPC) method (an algorithm to compute feedback solutions of optimal control problems). One of the main achievements of this doctoral work was showing that convergence of the MPC method can be proved by means of the *metric subregularity property* (to be introduced in the next section). Finally, we will comment on optimal control problems constrained by elliptic partial differential equations, and the pertaining results in this thesis. The contributions of the author to each paper are described at the beginning of each chapter.

1.1 Metric regularity properties

Metric regularity has become one of the central concepts of variational analysis in the last 2-3 decades. The roots of this concept go back to a set of fundamental notions of regularity in classical analysis, which are reflected in results such as the Implicit Function Theorem, Banach's Open Mapping Theorem, the theorems of Lyusternik and Graves, and Sard's Theorem; see [22] for a broader discussion.

1.1.1 Motivation

In order to motivate the subsequent definitions, let as begin talking about optimization. A general *optimization problem* can be described as a search problematic; one looks for an element of a set $\mathcal{C}$ that minimizes (or maximizes[1]) a real-valued function $f : \mathcal{C} \to \mathbb{R}$. More precisely, an element $x^* \in \mathcal{C}$ minimizes f over $\mathcal{C}$ if

$$f(x^*) \leq f(x) \quad \text{for every } x \in \mathcal{C}.$$

This problem is usually written as

$$\min_{x\in\mathcal{C}} f(x). \tag{1.1}$$

The function f is sometimes called the objective, the cost or the performance index, depending on the circumstance and the nature of the problem. Let us get now into more technical details. In the presence of a topology on $\mathcal{C}$, one can talk about *local minima*. An element $x^* \in \mathcal{C}$ si a local minimum of f over $\mathcal{C}$ if there exists a neighborhood $\mathcal{O}$ of x^* such that

$$f(x^*) \leq f(x) \quad \text{for every } x \in \mathcal{O} \subset \mathcal{C}.$$

Suppose now that $\mathcal{C}$ is a convex subset of a normed linear space X, and that $\mathcal{C}$ is endowed with the relative topology induced by the norm on X. If the directional derivatives of f exist, then the first order necessary condition for a local minimizer $\bar{x} \in \mathcal{C}$ is

$$\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \forall x \in \mathcal{C}. \tag{1.2}$$

This can be readily seen from the definition of directional derivative. Let X^* denote the topological dual of X. Defining

$$N_{\mathcal{C}}(x) := \{z \in X^* : \langle z, y - x \rangle \leq 0 \quad \forall y \in \mathcal{C}\},$$

[1]We restrict ourselves to talk only about minimization problems. It is well known that such problems can be stated in an equivalent way as minimization ones, see, e.g., [16, p. 11]

we can rewrite the first order necessary condition as the inclusion (generalized equation)

$$0 \in \nabla f(\bar{x}) + N_{\mathcal{C}}(\bar{x}). \tag{1.3}$$

Up to this point, it might seem that going from (1.2) to (1.3) is a mere phenomenon of semantics; but as we will soon see, it leads to some interpretations and results that can be later applied to problem (1.1). For example, consider the family of *perturbed problems* $\{\mathcal{P}_\xi\}_{\xi \in X^*}$ given by

$$\min_{x \in \mathcal{C}} f(x) - \xi x. \tag{1.4}$$

Given $\xi \in X^*$, the first order necessary condition for a minimizer $x \in \mathcal{C}$ of $\mathcal{P}_\xi$ is

$$\xi \in \nabla f(x) + N_{\mathcal{C}}(x). \tag{1.5}$$

Thus, the behavior of the solutions of each problem $\mathcal{P}_\xi$ can be analyzed by means of the *perturbed inclusions* (1.5). In the next subsection, we will review some of the basic definitions in metric regularity theory, this will enable us to give better a better description of dependence of solutions of (1.5) on the perturbations.

One of the most important cases of generalized equations are the ones induced by differential variational inequalities, see [2, 9]. These variational inequalities pop up naturally in optimal control. A broad discussion of the importance of these variational inequalities as models of optimization problems can be found in [28]. These are objects of the type

$$\begin{cases} 0 & = & Dy - g(x,y) \\ 0 & \in & h(x,y) + N_{\mathcal{C}}(y), \end{cases}$$

where D means differentiation in a space of functions $\mathcal{Y}$. Such a differential variational inequality, accompanied with side conditions for $\mathcal{Y}$, typically arises from the optimality conditions in optimal control theory.

1.1.2 Some definitions of metric regularity

In this subsection, we review some of the basic definitions that will be used in subsequent sections. Let (M_1, d_{M_1}) and (M_2, d_{M_2}) be metric spaces, and $\Phi : M_1 \rightrightarrows M_2$ be a set-valued mapping.

Consider the generalized equation

$$0 \in \Phi(p). \tag{1.6}$$

A solution of (1.6) is an element $\bar{p} \in M_1$ such that $0 \in \Phi(\bar{p})$. Given a metric space (M, d_M), the closed ball centered at $p \in M$ of radius $\alpha > 0$ is denoted by $\mathbb{B}_M(p, \alpha)$.

Definition 1.1.1. Let $\bar{p} \in M_1$ with $0 \in \Phi(\bar{p})$. The mapping Φ is said to be strongly (metrically) regular at $\bar{p}$ if there exist positive numbers α, β and κ such that for each $q \in \mathbb{B}_{M_2}(0, \beta)$ there exists a unique $p(q) \in \mathbb{B}_{M_1}(\bar{p}, \alpha)$ satisfying $q \in \Phi\big(p(q)\big)$, and

$$d_{M_1}\big(p(q_1), p(q_2)\big) \leq \kappa d_{M_2}(q_1, q_2)$$

for all $q_1, q_2 \in \mathbb{B}_{M_2}(0, \beta)$.

One of the simplest cases of strong regularity is that of a linear operator from a euclidean space onto itself; it is strongly regular if, and only if, its associated matrix is invertible. The notion in Definition 1.1.1 was introduced in [37] by Robinson in the 80's, see also [18, Section 3G].

The strong regularity property has many useful applications. One of the most popular is the analysis of convergence of the Newton method for generalized equations; see [18, Section 6C]. This analysis relies on an adequate contraction mapping principle for set-valued mappings, see [18, 5E.3], which in turn relies on the fact that the strong regularity property is stable with respect to adequate perturbations (Lipschitz functions); this is the so-called Lyusternik-Graves Theorem, see e.g., [18, Theorem 2B.1] or [15, Theorem 1].

Nevertheless, the metric regularity property is not enough to cope with some of the applications, hence the following definition was introduced in [36, Introduction], where a detailed explanation of the reasons for appropriateness of this concept is given. Let E_2 be a subset of M_2 and d_{E_2} a metric in E_2 such that (E_2, d_{E_2}) is continuously embedded in (M_2, d_{M_2}) and $d_{M_2} \leq c d_{E_2}$ for some $c > 0$.

Definition 1.1.2. Let $\bar{p} \in M_1$ with $0 \in \Phi(\bar{p})$. The mapping Φ is said to be strongly bi-metrically regular at $\bar{p}$ if there exist positive numbers α, β and κ such that for each $q \in \mathbb{B}_{E_2}(0,\beta)$ there exists a unique $p(q) \in \mathbb{B}_{M_1}(\bar{p},\alpha)$ satisfying $q \in \Phi\big(p(q)\big)$, and

$$d_{M_1}\big(p(q_1), p(q_2)\big) \leq \kappa d_{M_2}(q_1, q_2)$$

for all $q_1, q_2 \in \mathbb{B}_{E_2}(0,\beta)$.

A Lyusternik-Graves type theorem for the bi-metric regularity property was proved in [32], where also a study of the Newton method with this property was carried out, involving at each step linear quadratic optimal control problems.

One can weaken the concept of regularity to include more features, such as two metrics in the domain or Hölder estimates instead of Lipschitz ones, etc. We present now an extension of the definition of metric subregularity, see [1, 10].

Let E_1 be a subset of M_1 and d_{E_1} a metric in E_1 such that (E_1, d_{E_1}) is continuously embedded in (M_1, d_{M_1}) and $d_{E_1} \leq c d_{M_1}$ for some $c > 0$.

Definition 1.1.3. Let $\bar{p} \in E_1$ with $0 \in \Phi(\bar{p})$. The mapping Φ is said to be strongly (metrically) ι-subregular at $\bar{p}$ if there exist positive numbers α and κ such that

$$d_{E_1}(p, \bar{p}) \leq \kappa d_{M_2}(q, 0)^{\iota} \tag{1.7}$$

for all $q \in M_2$ and $p \in E_1 \cap \mathbb{B}_{M_1}(p,\alpha)$ satisfying $q \in \Phi(p)$. If $\iota = 1$, we simply say that Φ is strongly subregular at $\bar{p}$.

Results in the spirit of Lyusternik-Graves Theorem and analysis of numerical methods, such as the Newton one, have also been studied in the context of the subregularity property; see [10].

1.2 Optimal control problems for ODE's

In this subsection, we introduce an optimal control problem in Lagrange form. This is not a restrictions since the general Bolza problem can be transformed into a Lagrange one. We talk briefly on the notation before introducing the control problem. The elements of $\mathbb{R}^n$ are seen as column vectors. The inner products and the norms in the euclidean spaces are denoted by $\langle \cdot, \cdot \rangle$ and $|\cdot|$, respectively. Given a function $\psi : \mathbb{R}^s \to \mathbb{R}^r$ of variable z, we denote by $\psi_z(z)$ its derivative at a point $z \in \mathbb{R}^s$, represented by an $r \times s$-matrix. For a function $\psi : \mathbb{R}^{s+q} \to \mathbb{R}$ of the variables z and v, $\psi_{zv}(z,v)$ denotes its mixed second derivative at a point (z,v), represented by a $(s \times q)$-matrix. The space $L^k\left([a,b];\mathbb{R}^r\right)$, $k \in \{1,\infty\}$, of functions $\psi : [a,b] \to \mathbb{R}^r$ has the usual meaning. As usual, $W^{1,k}([a,b];\mathbb{R}^r)$ denotes the space of absolutely continuous functions $\psi : [a,b] \to \mathbb{R}^r$ for which the first derivative belongs to $L^k([a,b];\mathbb{R}^r)$. We abbreviate sometimes $W^{1,1} := W^{1,1}([0,T];\mathbb{R}^n)$, $L^1 := L^1\left([0,T];\mathbb{R}^m\right)$.

1.2.1 The control model

Let $T > 0$ and $y_0 \in \mathbb{R}^n$. The space $\mathcal{Y} := \big\{y \in W^{1,\infty}\big(0,T;\mathbb{R}^n\big) : y(0) = y_0\big\}$ is interpreted as the state space, and $\mathcal{U} \subset L^1\big(0,T;\mathbb{R}^m\big)$ as the control set. Suppose that for each $u \in \mathcal{U}$ there exists a unique function $y_u \in \mathcal{Y}$ such that

$$\dot{y}_u = f(\cdot, y_u, u), \tag{1.8}$$

where $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is a given function. Equation (1.8) represents the evolution of the system in time. The optimal control problem is the following:

$$\min_{u \in \mathcal{U}} \Big\{ \int_0^T g(t, y_u(t), u(t))\, dt \Big\}, \tag{1.9}$$

where $g : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a given real-valued function. We make the following assumption for the optimal control problem (1.8)-(1.9).

Assumption 1.2.1. *The following statements hold.*

(a) *The functions f and g are twice differentiable, and these functions, their first derivatives and their second derivatives are Lipschitz.*

(b) *There exists a closed convex set $U \subset \mathbb{R}^m$ such that*

$$\mathcal{U} = \left\{u \in L^\infty\big(0,T;\mathbb{R}^m\big) : u(t) \in U \quad \text{for a.e. } t \in [0,T]\right\}.$$

Assumption 1.2.1 ensures that the subsequent mathematical objects are well defined. We define the objective functional $\mathcal{J} : \mathcal{U} \to \mathbb{R}$ by

$$\mathcal{J}(u) := \int_0^T g(t, y_u(t), u(t))\, dt.$$

Definition 1.2.2. We say that $\bar{u} \in \mathcal{U}$ is a weak local minimizer of the optimal control problem (1.8)-(1.9) if there exists $\varepsilon_0 > 0$ such that $\mathcal{J}(\bar{u}) \le \mathcal{J}(u)$ for all $u \in \mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u}, \varepsilon_0)$.

We define the costate space by

$$\Lambda := \left\{\lambda \in W^{1,\infty}\big(0,T;\mathbb{R}^n\big) : \lambda(T) = 0\right\},$$

and the Hamiltonian $H : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ by

$$H(t, y, \lambda, u) := g(t, y, u) + \langle \lambda, f(t, y, u)\rangle.$$

The first order necessary conditions can be summarized in the next theorem.

Theorem 1.2.3. *Let Assumption 1.2.1 hold. If $\bar{u} \in \mathcal{U}$ is a local minimizer of problem (1.8)-(1.9), then there exists a unique $\lambda_{\bar{u}} \in \Lambda$ such that*

$$\begin{cases} 0 &= -\dot{y}_{\bar{u}} + f(\cdot, y_{\bar{u}}, \bar{u}), \\ 0 &= \dot{\lambda}_{\bar{u}} + H_y(\cdot, y_{\bar{u}}, \lambda_{\bar{u}}, \bar{u})^\top \\ 0 &\in H_u(\cdot, y_{\bar{u}}, \lambda_{\bar{u}}, \bar{u})^\top + N_{\mathcal{U}}(\bar{u}), \end{cases} \tag{1.10}$$

where the (restricted) normal cone at u to the set $\mathcal{U}$ is given by

$$N_{\mathcal{U}}(u) = \left\{\sigma \in L^\infty\big(0,T;\mathbb{R}^m\big) : \int_0^T \langle \sigma(t), w(t) - u(t)\rangle\, dt \le 0 \quad \forall w \in \mathcal{U}\right\}.$$

We now proceed to re-state the system of first order necessary conditions (1.10) as generalized equation. Consider the sets $\mathcal{M}_r := \mathcal{Y} \times \Lambda \times \mathcal{U}$, and

$$\mathcal{Z}_r := L^r\big(0,T;\mathbb{R}^n\big) \times L^r\big(0,T;\mathbb{R}^n\big) \times L^\infty\big(0,T;\mathbb{R}^m\big), \quad r \in \{1,2\}. \tag{1.11}$$

Let $d_{\mathcal{M}_r}$ and $d_{\mathcal{Z}_r}$ be the shift-invariant metrics in $\mathcal{M}_r$ and $\mathcal{Z}_r$, respectively, given by

$$d_{\mathcal{M}_r}(y,p,u) := |y|_{W^{1,r}} + |\lambda|_{W^{1,r}} + |u|_{L^r},$$
$$d_{\mathcal{Z}_r}(\xi,\eta,\rho) := |\xi|_{L^r} + |\eta|_{L^r} + |\rho|_{L^s},$$

where $1/r + 1/s = 1$. The optimality mapping is defined as the set-valued mapping $\Phi : \mathcal{M}_r \twoheadrightarrow \mathcal{Z}_r$ given by

$$\Phi(y,\lambda,u) = \begin{pmatrix} -\dot{y} + f(\cdot, y, u) \\ \dot{\lambda} + H_y(\cdot, y, \lambda, u)^\top \\ H_u(\cdot, y, \lambda, u)^\top + N_{\mathcal{U}}(u) \end{pmatrix}. \tag{1.12}$$

Then the optimality system (1.10) can be recast as the inclusion

$$0 \in \Phi(y, \lambda, u). \tag{1.13}$$

We can now employ the notions of regularity introduced in Section 1.1 to study the stability of system (1.10) in terms of inclusion (1.13).

The metric regularity of the optimality mapping has been studied under the classic (integral) coercivity condition.

Coercivity condition. There is a constant $c_0 > 0$ such for any $v \in \mathcal{U} - \mathcal{U}$ the inequality

$$\int_0^T [\langle H_{yy}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) z_v, z_v\rangle + 2\langle H_{uy}(\cdot, y_{\bar{u}}, p_{\bar{u}}) z_v, v\rangle + \langle H_{uu}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) v, v\rangle]\, dt \geq c_0 \|v\|^2_{L^2}$$

is fulfilled, where z_v is the unique solution of the equation $\dot{z}_v = f_y(\cdot, y_{\bar{u}}, \bar{u}) z_v + f_u(\cdot, y_{\bar{u}}, \bar{u}) v$ with $z_v(0) = 0$.

This assumption is standard in optimal control; it goes back to the work of Hager [20], where it was used to show convergence of the multiplier method in optimal control. It was also used in [17] to prove results about convergence of discrete approximations in optimal control. One can interpret coercivity as a strong second order sufficient condition.

Theorem 1.2.4. *[16, Theorem 17.2] Suppose that the coercivity condition holds. Then the optimality mapping is strongly regular at $(y_{\bar{u}}, \lambda_{\bar{u}}, \bar{u})$ as mapping from $\mathcal{M}_2$ to $\mathcal{Z}_2$.*

There many useful applications of Theorem 1.2.4, see e.g., [17, Section 7] where some convergence results of the Euler method are obtained.

1.2.2 Results on affine optimal control problems

We say that problem (1.8)-(1.9) is affine if the functions f and g have the following form

$$f(t, x, u) = a(x) + B(x)u \quad \text{and} \quad g(t, x, u) = w(x) + s(x)u. \tag{1.14}$$

The case of affine problems is especially challenging due to the typical discontinuity of the optimal control functions. In fact, assumptions like the coercivity condition cannot hold for this type of problems.

This impossibility is because the coercivity condition implies the Legendre-Clebsch condition (a.e. pointwise coercivity condition), see [18].

Legendre-Clebsch condition. There is a constant $c_0 > 0$ such for any $v \in U - U$ and for almost every $t \in [0, T]$ the inequality

$$\langle H_{uu}(t, y_{\bar{u}}(t), p_{\bar{u}}(t), \bar{u}(t)) v(t), v(t)\rangle \geq c_0 \|v(t)\|^2$$

is fulfilled.

From the latter condition we see that if $H_{uu}(t, y_{\bar{u}}(t), p_{\bar{u}}(t) = 0$ a.e in $[0, T]$ then the coercivity condition cannot hold (unless U is a singleton).

Before stating our contributions, let us comment on some of the results in the literature. A general sufficient condition for strong sub-regularity of the mapping Φ in the single metric $d_{\mathcal{M}_1}$ in $\mathcal{M}_1$ is given in [27, Theorem 3.1]. It involves the following assumption at a local minimizer $\bar{u} \in \mathcal{U}$.

Assumption. There is a constant $c_0 > 0$ such for any $v \in \mathcal{U} - \bar{u}$ the inequality

$$\int_0^T \langle H_u(\cdot, y_{\bar{u}}, p_{\bar{u}}), v\rangle\, dt + \int_0^T [\langle H_{yy}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) z_v, z_v\rangle + 2\langle H_{uy}(\cdot, y_{\bar{u}}, p_{\bar{u}}) z_v, v\rangle]\, dt \geq c_0 \|v\|^2_{L^1}$$

is fulfilled, where z_v is the unique solution of the equation $\dot{z}_v = f_y(\cdot, y_{\bar{u}}, \bar{u}) z_v + f_u(\cdot, y_{\bar{u}}, \bar{u}) v$ with $z_v(0) = 0$.

Consider the metric space given by

$$\mathcal{Z}^* := L^1\big(0, T; \mathbb{R}^n\big) \times L^1\big(0, T; \mathbb{R}^n\big) \times W^{1,\infty}\big(0, T; \mathbb{R}^m\big),$$

endowed with the shift invariant metric

$$d_{\mathcal{Z}^*}(\xi, \eta, \rho) := |\xi|_{L^1} + |\eta|_{L^1} + |\rho|_{W^{1,\infty}}.$$

However, for affine problems, the strong bi-metric regularity property is more suitable, see [36, 35]. In [36, Corollary 3.5], sufficient conditions for strong bimetric regularity (with respect to the metric spaces $\mathcal{M}_1, \mathcal{Z}_1$ and $\mathcal{Z}^*$) for a Mayer problem were given. These conditions involve the non-negativity of the second variation of the objective functional and a growth condition at the optimal control.

Our results on affine optimal control problems are as follows. In [14, Theorem 2.1], we obtained more general sufficient conditions for the optimality mapping to be strongly bi-metrically regular (with respect to the metric spaces $\mathcal{M}_1, \mathcal{Z}_1$ and $\mathcal{Z}^*$) at an optimal control $\bar{u} \in \mathcal{U}$. One of the main features of these conditions is that they do not involve the second derivative of the associated Hamiltonian with respect to the control. Moreover, they do not require non-negativity of the second variation of the objective functional. We showed in [14, Section 4] an example where the assumptions are satisfied for a non-convex problem; this is a new finding in the optimal control context, in general. The assumptions in [14, Theorem 2.1] for an optimal control $\bar{u} \in \mathcal{U}$ can be read as follows, see [14, Assumption A2].

Assumption 1.2.5. *The matrix*

$$H_{uy}(t, y_{\bar{u}}(t), p_{\bar{u}}(t)) f_u(t, y_{\bar{u}}(t)) \quad \text{is symmetric for a.e. } t \in [0, T] \ .$$

Assumption 1.2.6. *There exist positive numbers α_0, γ_0 and c_0 such for any $v \in \mathcal{U} - \mathcal{U}$, and $\sigma \in \mathbb{B}_{W^{1,\infty}(\bar{\sigma},\gamma_0)} \cap (-N_{\mathcal{U}}(u))$, the inequality*

$$\int_0^T \langle \sigma, v \rangle \, dt + \int_0^T \left[\langle H_{yy}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) z_v, z_v \rangle + 2 \langle H_{uy}(\cdot, y_{\bar{u}}, p_{\bar{u}}) z_v, v \rangle \right] dt \geq c_0 \|v\|_{L^1}^2$$

is fulfilled. Here z_v is the unique solution of the equation $\dot{z}_v = f_y(\cdot, y_{\bar{u}}, \bar{u}) z_v + f_u(\cdot, y_{\bar{u}}, \bar{u}) v$ with $z_v(0) = 0$, and $\bar{\sigma} = H_u(\cdot, y_{\bar{u}}, \lambda_{\bar{u}})$.

Theorem 1.2.7. *[14, Theorem 2.1] Suppose that Assumptions 1.2.5 and 1.2.6 hold. Then the optimality mapping Φ in (1.12) is strongly bi-metrically regular at $(y_{\bar{u}}, \lambda_{\bar{u}}, \bar{u})$ with respect to the metric spaces $\mathcal{M}_1, \mathcal{Z}_1$ and $\mathcal{Z}^*$.*

As an application, in [14, Section 5] we prove that the obtained sufficient conditions imply uniform first order convergence of the Euler discretization scheme when applied to affine problems that are close enough to a reference one.

1.3 Model predictive control method

The Model Predictive Control (MPC) is a method for on-line feedback approximation, the idea behind it can be described as follows. Observe/measure the current state of a system, and from that compute very rapidly an open-loop control solution to the problem at hand. The first portion of this function is then used during a short time interval, after that a new measurement of the state is made. Repeating this over a time interval will yield a feedback controller synthesis from knowledge of open-loop controls obtained, see [24, 38].

The main advantages of the MPC paradigm are that the feedback nature of the method provides additional robustness to disturbances, modeling inaccuracies, and implementation errors, in contrast to classical optimal control theory. The MPC is a powerful method, but the rigorous mathematical theory investigating the scope of validity and the efficiency of the MPC method under appropriate suppositions is still underdeveloped.

This section contains a brief summary of the results in [1] including a short subsection with some of the technical details of the problem as well as a description of the MPC algorithm.

1.3.1 The model

Consider now the following optimal control problem, further denoted by $\mathcal{P}_\pi(0, y_0)$:

$$\min_{u \in \mathcal{U}} \left\{ \mathcal{J}_\pi(u) := \int_0^T g(\pi(t), y(t), u(t)) \, dt \right\}, \tag{1.15}$$

subject to

$$\dot{y}(t) = f(\pi, y, u) \quad y(0) = y_0. \tag{1.16}$$

Here the state vector $y(t)$ belongs to $\mathbb{R}^n$ for each $t \in [0,T]$, and the control function $u(\cdot)$ belongs to the set $\mathcal{U}$ of all Lebesgue measurable functions $u : [0,T] \to U$, where $U \subset \mathbb{R}^m$. The function π represents an uncertain time-dependent parameter which is known to belong to a set Π of bounded Lebesgue measurable functions $\pi : [0,T] \to \mathbb{R}^l$. Correspondingly, f and g are defined on $\mathbb{R}^l \times \mathbb{R}^n \times \mathbb{R}^m$ with values in $\mathbb{R}^n$ and $\mathbb{R}$, respectively. The initial state $y_0 \in \mathbb{R}^n$ and the final time $T > 0$ are fixed.

Problem $\mathcal{P}_\pi(0, y_0)$ will be considered under the following assumptions.

Assumption 1.3.1. *The set U is convex and compact. The functions f and g are two times differentiable with respect to (y,u), these functions and their first and second derivatives in (y,u) are Lipschitz continuous with respect to (π, y, u).*

For any $\pi \in \Pi$, along with problem $\mathcal{P}_\pi(0, y_0)$ we consider the family, denoted by $\mathcal{P}_\pi(\tau, y_\tau)$, consisting of problems which have the same form as (1.15)–(1.16) but with the initial time 0 replaced with any $\tau \in [0,T)$ and y_0 replaced with any $y_\tau \in \mathbb{R}^n$. Of course, then only the restriction of the parameter ξ to $[\tau, T]$ matters.

Assumption 1.3.2. *For every $u \in \mathcal{U}$, $y_0 \in \mathbb{R}^n$, and $\pi \in \Pi$ equation (1.16) has a solution y on $[0,T]$ (which is then unique due to Assumption 1.3.1). For every $\tau \in [0,T)$, $y_\tau \in \mathbb{R}^n$ and $\pi \in \Pi$ problem $\mathcal{P}_\pi(\tau, y_\tau)$ has an optimal solution.*

Optimality in the last assumption means local optimality of the objective functional with respect to the L^1-norm of the controls. We now reformulate the optimality system in functional spaces. The space $L^q(\tau,T)$, $q = 1, 2, \ldots, \infty$, of vector functions on $[\tau, T]$ (with any fixed dimension) has the usual meaning, with the norm denoted by $|\cdot|_{L^q}$. The space of all absolutely continuous vector functions on $[\tau, T]$ is denoted by $W^{1,1}(\tau, T)$, with the norm $|y|_{W^{1,1}} = |y|_{L^1} + |\dot{y}|_{L^1}$. Moreover, $W^{1,1}_T(\tau,T)$ is the space of functions $\lambda \in W^{1,1}(\tau,T)$ for which $\lambda(T) = 0$. The notations of norms do not include the time horizon, but it will be clear from the context. For the same reason we often skip the time horizon from the notations of spaces.

Denote

$$\mathcal{M}_\tau := W^{1,1}(\tau,T) \times W^{1,1}_T(\tau,T) \times \mathcal{U}_\tau, \qquad \mathcal{Z}_\tau := L^1(\tau,T) \times \mathbb{R}^n \times L^1(\tau,T) \times L^\infty(\tau,T),$$

where $\mathcal{U}_\tau = \{u \in L^1(\tau,T) : u(t) \in U \text{ for a.e. } t \in [\tau,T]\}$ is the set of admissible control functions on $[\tau, T]$, thus $\mathcal{U}_0 = \mathcal{U}$. We also set $\mathcal{M} := \mathcal{M}_0$ and $\mathcal{Z} := \mathcal{Z}_0$. The metrics in $\mathcal{M}_\tau$ and $\mathcal{Z}_\tau$ are given in terms of norms as follows: for $p = (y, \lambda, u) \in \mathcal{M}_\tau$ and $q = (\xi, \nu, \eta, \rho) \in Z_\tau$

$$d_{\mathcal{M}_\tau}(p,0) := |p|_{\mathcal{M}_\tau} := |y|_{W^{1,1}} + |\lambda|_{W^{1,1}} + |u|_{L^1}, \quad d_{\mathcal{Z}_\tau}(q,0) := |q|_{\mathcal{Z}_\tau} := |\xi|_{L^1} + |\nu|_{\mathbb{R}^n} + |\eta|_{L^1} + |\rho|_{L^\infty}.$$

In addition, we define in $\mathcal{M}^*_\tau$ as follows. Let $\Gamma \subset [0,T]$ be a fixed finite set. For $u_1, u_2 \in \mathcal{U}_\tau$ denote

$$d^*(u_1,u_2) := \inf\{\varepsilon > 0 : |u_1(t) - u_2(t)| \le \varepsilon \text{ for a.e. } t \in [0,T] \setminus (\Gamma + [-\varepsilon, \varepsilon])\}. \tag{1.17}$$

Abusing the notation, we define $\mathcal{M}^*_\tau := \mathcal{M}_\tau$ with the shift-invariant metric

$$d_{\mathcal{M}^*_\tau}(p,0) := |y|_{W^{1,1}} + |\lambda|_{W^{1,1}} + d^*(u,0).$$

For any $\pi \in \Pi$, $\tau \in [0,T)$, and $y_\tau \in \mathbb{R}^n$ define on $\mathcal{M}_\tau$ the *optimality map* by

$$\Phi_{(\pi,\tau,y_\tau)}(y) = \begin{pmatrix} -\dot{y} + f(p,y,u) \\ y(\tau) - y_\tau \\ \dot{\lambda} + H_y(\pi,y,\lambda,u)^\top \\ H_u(\pi,y,\lambda,u)^\top + N_{\mathcal{U}_\tau}(u) \end{pmatrix},$$

where now $N_{\mathcal{U}_\tau}(u)$ is the normal cone to $\mathcal{U}_\tau$ at u in the space $L^1(\tau,T)$, that is,

$$N_{\mathcal{U}_\tau}(u) = \{\sigma \in L^\infty(\tau,T) : \int_\tau^T \langle \sigma, w - u \rangle \text{ for all } w \in \mathcal{U}\}.$$

With these notations one can recast the optimality system for problem $\mathcal{P}_\pi(\tau, y_\tau)$ as

$$0 \in \Phi_{(\pi,\tau,y_\tau)}(y, \lambda, u).$$

Obviously, due to the compactness of the set U, $\Phi_{(\pi,\tau,y_\tau)}$ is a set-valued mapping from $\mathcal{M}_\tau$ to $\mathcal{Z}_\tau$.

1.3.2 The algorithm and the results

Let us fix a reference parameter $\bar{\pi} \in \Pi$ and denote by $(\bar{y}, \bar{u})$ a solution of problem $\mathcal{P}_{\bar{\pi}}(0, y_0)$ (see Remark **??**). Let $\bar{\lambda}$ be the corresponding adjoint function, so that the $0 \in \Phi(\bar{y}, \bar{\lambda}, \bar{u})$. The following assumption plays a key role in the error analysis of the MPC method presented in the next section.

Assumption 1.3.3. *The map $\Phi_{(\bar{p},0,x_0)} : \mathcal{M} \twoheadrightarrow \mathcal{Z}$ is strongly sub-regular at $(\bar{y}, 0)$ (in the metrics $d_{\mathcal{M}}$ and $d_{\mathcal{M}^*}$).*

The finite set Γ that appears in the definition of the metric d^* is arbitrary, but it is appropriately specified in [1, Section 4] for several classes of problems, together with sufficient conditions for Assumption 1.3.3.

It is assumed that for some parameter $\bar{\pi} \in \Pi$ equation (1.16) with $\pi = \bar{\pi}$ reproduces a "real" system, the states of which can be measured (with a measurement errors). Recall that $(\bar{y}, \bar{u})$ is a reference optimal solution of $\mathcal{P}_{\bar{\pi}}(0, y_0)$. We describe now the MPC algorithm.

Given a natural number N, we denote by $\{t_k\}_{k=0}^N$ the grid with step-size $h = T/N$, that is, $t_k = kh$, $k = 0, \ldots, N$. To describe the k-th stage of the MPC algorithm we assume that an admissible control function u^N is already determined on $[0, t_k]$ and applied to the "real" system. Denote by y^N the corresponding trajectory, that is the solution of (1.16) with $\pi = \bar{\pi}$ and $u = u^N$. Then the state of the "real" system is measured with a measurement error e_k, that is, the vector $y_k^0 = y^N(t_k) + e_k$ becomes available at time t_k. In addition, a prediction $\pi_k \in \Pi$ for the time horizon $[t_k, T]$ is made. Then an approximate solution $(\tilde{y}_k, \tilde{u}_k) \in W^{1,1} \times \mathcal{U}_{t_k}$ of the problem $\mathcal{P}_{\pi_k}(t_k, y_k^0)$ is found, and u^N is extended to $[0, t_{k+1}]$ as $u^N(t) = \tilde{u}_k(t)$ for $t \in (t_k, t_{k+1}]$.

The process continues in the same way as long as $k < N$. The control u^N is called *MPC-generated control* and the corresponding trajectory x^N of the "real" system (1.16) with $u = u^N$ and $\pi = \bar{\pi}$ is called *MPC-generated trajectory.*

The quality of a prediction $\pi_k \in \Pi$ on $[t_k, T]$ will be measured by the norm $e_k^\pi := \|\pi_k - \bar{\pi}_{[t_k,T]}\|_{L^\infty}$, and the pair $(\tilde{y}_k, \tilde{u}_k)$ is an approximate solution of problem $\mathcal{P}_{\pi_k}(t_k, y_k^0)$ in the sense that for some absolutely continuous $\tilde{\lambda}_k$ the triplet $\tilde{p}_k := (\tilde{y}_k, \tilde{\lambda}_k, \tilde{u}_k)$ satisfies the inclusion (approximate optimality conditions)

$$\tilde{q}_k \in \Phi_{(\pi_k, t_k, y_k^0)}(\tilde{p}_k) \tag{1.18}$$

with some $\tilde{q}_k \in \mathcal{Z}_{t_k}$. Most of the numerical methods for optimal control give approximations with a small residual $\tilde{q}_k$. The norm $e_k^u := |\tilde{q}_k|_{\mathcal{Z}_{t_k}}$ of the residual will be used as a measure of the accuracy of the approximate solution $(\tilde{y}_k, \tilde{u}_k)$ of problem $\mathcal{P}_{\pi_k}(t_k, y_k^0)$.

The main result in [1] gives an estimate of the difference between the MPC-generated control and the optimal open-loop control for the reference problem (the latter corresponding to the "ideal" scenario where the prediction, the measurement, and the solution of the auxiliary problems are all exact). A remarkable feature of the estimation is that the overall error of the MPC-generate control depends on the *average* of the errors appearing at the steps of the algorithm, thus occasional relatively large errors in the prediction or measurement do not substantially damage the MPC-generated control. Another interesting feature of the overall error is that for some classes of problems it depends linearly on the averaged errors appearing at the steps of the method, while for other classes, the estimate of the overall error depends on the square root of the averaged errors (and this estimate is sharp, see [1, Example 3.5]).

Theorem 1.3.4. *[1, Theorem 3.1] Let Assumptions (A1)–(A3) be fulfilled. Then there exists numbers N_0, $\delta > 0$, C_1, C_2, and C_3 such that for any natural number $N \geq N_0$, for any sequence of measurement errors $\{e_k\}$, for any sequence of predictions $\pi_k \in \Pi$ and approximation errors $\{e_k^u\}$ satisfying the conditions*

$$|e_k| + e_k^p + e_k^u \leq \delta, \qquad \|\tilde{u}_k - \bar{u}\|_{L^1} \leq \delta, \quad k = 0, \ldots, N-1,$$

any MPC-generated trajectory-control pair (y^N, u^N) satisfies the estimate

$$\|u^N - \bar{u}\|_1 + \|y^N - \bar{y}\|_{1,1} \leq \begin{cases} C_1 \mathcal{E} & \text{if } \Gamma = \emptyset, \\ C_2\sqrt{\mathcal{E}} + C_3 h & \text{if } \Gamma \neq \emptyset, \end{cases}$$

where

$$\mathcal{E} := \frac{1}{N} \sum_{k=0}^{N-1} (|e_k| + e_k^\pi + e_k^u)$$

is the averaged error appearing at the MPC steps.

1.4 Elliptic optimal control problems

The optimal control of ordinary differential equations cannot cover all interesting and practical applications, for instance, heat conduction, electromagnetic waves, fluid flows, freezing processes, etc. In this section we introduce an optimal control problem governed by elliptic partial differential equations. Later in the section, we comment the results obtained in the papers [12, 13]. The motivations for studying stability of solutions comes from the error analysis of numerical methods, see e.g., [29, 30].

1.4.1 The control model

Consider the following optimal control problem

$$\min_{u\in\mathcal{U}} \left\{ \int_\Omega L(x,y)\,dx \right\}, \tag{1.19}$$

subject to

$$\mathcal{L}y + d(\cdot,y) = u \tag{1.20}$$

Here, $\mathcal{L} : D(\mathcal{L}) \to L^2(\Omega)$ is an elliptic operator representing the action of $-\operatorname{div}(A\nabla)$ and $D(\mathcal{L})$ is a set encapsulating[2] the boundary conditions of a partial differential equation. More explicitly, (1.20) represents either the Robin boundary value problem

$$\begin{cases} -\operatorname{div}\big(A(x)\nabla y\big) + d(x,y) &= u \quad \text{in} \quad \Omega \\ A(x)\nabla y\cdot\nu + y &= 0 \quad \text{on} \quad \partial\Omega. \end{cases} \tag{1.21}$$

or the Dirichlet value problem

$$\begin{cases} -\operatorname{div}\big(A(x)\nabla y\big) + d(x,y) &= u \quad \text{in} \quad \Omega \\ y &= 0 \quad \text{on} \quad \partial\Omega. \end{cases} \tag{1.22}$$

The set $\Omega\subset\mathbb{R}^n$ is a bounded domain with Lipschitz boundary, where $n\in\{2,3\}$. The unit outward normal vector field on the boundary $\partial\Omega$, which is single valued a.e. in $\partial\Omega$, is denoted by ν. The control set is given by

$$\mathcal{U} := \{u:\Omega\to\mathbb{R} \text{ measurable}: \; b_1\le u(x)\le b_2 \text{ for a.e. } x\in\Omega\},$$

where b_1 and b_2 are positive numbers satisfying $b_1\le b_2$. The functions $L, d:\Omega\times\mathbb{R}\to\mathbb{R}$ are real-valued and measurable, and $A:\Omega\to\mathbb{R}^{n\times n}$ is a measurable matrix-valued function.

Assumption 1.4.1. *The following statements are assumed to hold.*

(i) The set $\Omega\subset\mathbb{R}^n$ is a bounded Lipschitz domain. The matrix $A(x)$ is symmetric for a.e. x in Ω, and there exists $\alpha>0$ such that $\xi\cdot A(x)\xi\ge\alpha|\xi|^2$ for a.e. x in Ω and all $\xi\in\mathbb{R}^n$.

(ii) The functions L and d are Carathéodory, twice differentiable with respect to the second variable, and their second derivatives are locally Lipschitz uniformly in the first variable.

(iii) The functions $A, L(\cdot,0), L_y(\cdot,0), d(\cdot,0)$ and $d_y(\cdot,0)$ are measurable and bounded.

(iv) The function $d_y(\cdot,y)$ is nonnegative a.e. in Ω for all $y\in\mathbb{R}$.

The necessary optimality conditions (Pontryagin principle) for problem (1.19)-(1.20) is given by the system

$$\begin{cases} 0 &= \mathcal{L}y + d(y) - u, \\ 0 &= \mathcal{L}\lambda + d_y(\cdot,y)\lambda - L_y(\cdot,y), \\ 0 &\in \lambda + N_{\mathcal{U}}(u), \end{cases} \tag{1.23}$$

[2]See [13, Section 2.1] for a precise description.

If $u \in \mathcal{U}$ is a local solution of problem (1.19)–(1.20), then the triple (y_u, λ_u, u) is a solution of (1.23). Let us recast system (1.23) in a functional frame. We introduce the metric spaces

$$\mathcal{M} := D(\mathcal{L}) \times D(\mathcal{L}) \times \mathcal{U} \quad \text{and} \quad \mathcal{Z} := L^2(\Omega) \times L^2(\Omega) \times L^\infty(\Omega),$$

endowed with the following metrics. For $\psi_i = (y_i, \lambda_i, u_i) \in \mathcal{M}$ and $\zeta_i = (\xi_i, \eta_i, \rho_i) \in \mathcal{Z}$, $i \in \{1, 2\}$,

$$d_{\mathcal{M}}(\psi_1, \psi_2) := |y_1 - y_2|_{L^2(\Omega)} + |\lambda_1 - \lambda_2|_{L^2(\Omega)} + |u_1 - u_2|_{L^1(\Omega)},$$
$$d_{\mathcal{Z}}(\zeta_1, \zeta_2) := |\xi_1 - \xi_2|_{L^2(\Omega)} + |\eta_1 - \eta_2|_{L^2(\Omega)} + |\rho_1 - \rho_2|_{L^\infty(\Omega)}.$$

Both metrics are shift-invariant. Then the optimality mapping is defined as the set-valued mapping $\Phi : \mathcal{M} \twoheadrightarrow \mathcal{Z}$ given by

$$\Phi(y, \lambda, u) = \begin{pmatrix} \mathcal{L}y + d(y) - u \\ L\lambda + d_y(\cdot, y)\lambda - L_y(\cdot, y) \\ \lambda + N_{\mathcal{U}}(u) \end{pmatrix}. \tag{1.24}$$

Then, as in previous sections, the optimality system (1.23) can be recast as the inclusion

$$0 \in \Phi(y, \lambda, u). \tag{1.25}$$

Our purpose is to study the stability of system (1.23), or equivalently of inclusion (1.25), with respect to perturbations in the right-hand side. We will consider the concept of subregularity, see Definition 1.1.3. From now on, we write $\bar\psi := (\bar y, \bar\lambda, \bar u) = (y_{\bar u}, \lambda_{\bar u}, \bar u)$, where $\bar u$ is a fixed local solution of problem (1.19)–(1.20).

The inequality (1.7) in the definition of ι-subregularity, can be written explicit for this problem as

$$|y - y_{\bar u}|_{L^2(\Omega)} + |\lambda - \lambda_{\bar u}|_{L^2(\Omega)} + |u - \bar u|_{L^1(\Omega)} \le \kappa \Big(|\xi|_{L^2(\Omega)} + |\eta|_{L^2(\Omega)} + |\rho|_{L^\infty(\Omega)} \Big)^{\iota}. \tag{1.26}$$

Hence, if the optimality mapping is ι-subregular, all solutions of the system

$$\begin{cases} \xi &= \mathcal{L}y + d(y) - u, \\ \eta &= L\lambda + d_y(\cdot, y)\lambda - L_y(\cdot, y), \\ \rho &\in \lambda + N_{\mathcal{U}}(u). \end{cases} \tag{1.27}$$

that are near $(y_{\bar u}, \lambda_{\bar u}, \bar u)$ satisfy the Hölder estimate (1.26) with respect to the perturbations $\zeta = (\xi, \eta, \rho)$, provided they are small enough.

1.4.2 The results on elliptic optimal control problems

Results for optimization problems constrained by partial differential equations have been gaining relevance in recent years, see [3, 6, 7, 8, 11, 33]. Most of the stability results for elliptic control problems are obtained under a second order growth condition (analogous to the classical Legendre-Clebsch condition). For literature concerning this type of problems, the reader is referred to [19, 21, 23, 25, 26, 34] and the references therein. However, its stability has been only investigated in a handful of papers, see e.g., [11, 31, 33].

Stability of controls

We begin reviewing one of relevant results in the literature. In [33], the authors consider linear perturbations in the state and adjoint equations. They use the so-called structural assumption (a growth assumption satisfied near the jumps of the control) on the adjoint variable. This assumption has been widely used in the literature on bang-bang control of ordinary differential equations in a somewhat different form.

In [13], we introduce in the literature the following assumption at a local minimizer $\bar u \in \mathcal{U}$.

Assumption 1.4.2. *There exist positive numbers α_0, γ_0 and $k^* \in [1, 4/n)$ such that*

$$\int_\Omega \lambda_{\bar u} v \, dx + \int_\Omega \left[L_{yy}(x, y_{\bar u}) - \lambda_{\bar u} d_{yy}(x, y_{\bar u}) \right] z_v^2 \, dt \ge \gamma_0 |v|_{L^1(\Omega)}^{k^*+1}, \tag{1.28}$$

for all $v \in \mathcal{U} - \bar u$ with $|v|_{L^1(\Omega)} \le \alpha_0$. Here z_v is the unique solution of the equation $\mathcal{L} z_v + d_y(\cdot, y_{\bar u}) z_v = v$.

In [13, Section 6], we prove that the last assumption is equivalent to

Assumption 1.4.3. *There exist positive numbers* α_0, γ_0 *and* $k^* \in [1, 4/n)$ *such that*

$$\int_\Omega \lambda_{\bar{u}} v \, dx + \int_\Omega \left[L_{yy}(x, y_{\bar{u}}) - \lambda_{\bar{u}} d_{yy}(x, y_{\bar{u}}) \right] z_v^2 \, dt \geq \gamma_0 |v|_{L^1(\Omega)}^{k^*+1},$$

for all $v \in \mathcal{U} - \bar{u}$ *with* $v \in C_{\bar{u}}^\tau \cap \mathbb{B}_{L^1(\Omega)}(\bar{u}; \alpha_0)$*, where* z_v *is the unique solution of the equation* $\mathcal{L} z_v + d_y(y_{\bar{u}}) z_v = v$.

For a fixed $\tau > 0$, the extended cone is given by

$$C_{\bar{u}}^\tau = \left\{ v \in L^2(\Omega) : v(x) \left\{ \begin{array}{lll} = 0 & \text{if} & |\lambda_{\bar{u}}(x)| > \tau \text{ or } \bar{u}(x) \in (b_1(x), b_2(x)) \\ \geq 0 & \text{if} & |\lambda_{\bar{u}}(x)| \leq \tau \text{ and } \bar{u}(x) = b_1(x) \\ \leq 0 & \text{if} & |\lambda_{\bar{u}}(x)| \leq \tau \text{ and } \bar{u}(x) = b_2(x) \end{array} \right. \right\}.$$

The equivalence between the assumptions shows that for $k^* \in [1, 4/n)$, Assumption 1.4.2 is weaker than the assumptions in [33], see [13, Section 6] for more details. Our main result in [13] is

Theorem 1.4.4. *Let Assumption 1.4.2 hold. Then the optimality mapping* Φ *is* ι*-subregular at* $\bar{\psi}$ *with* $\iota = 1/k^*$ *as a mapping from* $\mathcal{M}$ *to* $\mathcal{Z}$.

This result generalizes the one in [33, Theorem 4.5]. One of the main novelties of our work is that using Theorem 1.4.4, we proved stability of problem (1.19)-(1.20) with respect to nonlinear perturbations, see [13, Section 5] for precise details.

Stability of states

Results in state-stability were also proved in [12]; they involve weaker assumptions than in [13], but also yield less stronger theorems. There are some minor differences between the model in [12] and the one in [13], the main one is that in [12], we considered an elliptic PDE with Dirichlet boundary condition and with a convection term which makes the linear differential operator $\mathcal{L}$ not monotone nor coercive. Before summarizing our main result, we introduce new metric spaces

$$\mathcal{M}^* := \left\{ (y, \lambda) \in D(\mathcal{L}) \times D(\mathcal{L}) : 0 \in \lambda + N_{\mathcal{U}}\big(\mathcal{L} y + d(\cdot, y)\big) \right\} \quad \text{and} \quad \mathcal{Z}^* := L^2(\Omega) \times L^2(\Omega),$$

endowed with the following metrics. For $\psi_i = (y_i, \lambda_i) \in \mathcal{M}^*$ and $\zeta_i = (\xi_i, \eta_i) \in \mathcal{Z}^*$, $i \in \{1, 2\}$,

$$d_{\mathcal{M}^*}(\psi_1, \psi_2) := |y_1 - y_2|_{L^2(\Omega)} + |\lambda_1 - \lambda_2|_{L^2(\Omega)},$$
$$d_{\mathcal{Z}^*}(\zeta_1, \zeta_2) := |\xi_1 - \xi_2|_{L^2(\Omega)} + |\eta_1 - \eta_2|_{L^2(\Omega)}.$$

Consider also the optimality mapping $\Phi^* : \mathcal{M}^* \to \mathcal{Z}^*$ re-defined as the set-valued mapping given by

$$\Phi^*(y, \lambda) = \begin{pmatrix} \mathcal{L} y + d(y) - u \\ \mathcal{L} \lambda - L_y(\cdot, y) + \lambda d_y(\cdot, y) \end{pmatrix}. \tag{1.29}$$

Under a new assumption we proved Lipschitz stability of the associated states to the optimal controls with respect to perturbations in the equation and the objective functional, which can be translated as subregularity of the optimality mapping (1.29). To the best of our knowledge this is the first time this assumption is used in the literature, and is worth pointing out that this assumption pops up naturally studying the subregularity of the optimality mapping. The new assumption introduced to literature reads as follows.

Assumption 1.4.5. *There exist positive numbers* α_0, γ_0 *such that*

$$\int_\Omega \lambda_{\bar{u}} v \, dx + \int_\Omega \left[L_{yy}(x, y_{\bar{u}}) - \lambda_{\bar{u}} d_{yy}(x, y_{\bar{u}}) \right] z_v^2 \, dt \geq \gamma_0 |z_v|_{L^2(\Omega)}^2, \tag{1.30}$$

for all $v \in \mathcal{U} - \bar{u}$ *with* $|z_v|_{L^2(\Omega)} \leq \alpha_0$*. Here* z_v *is the unique solution of the equation* $\mathcal{L} z_v + d_y(\cdot, y_{\bar{u}}) z_v = v$.

We mention that the assumption is not only new in the context of optimal control of PDE's, but in general in optimal control. We are now ready to give the main result of [12] (stated in a equivalent different form in terms of the optimality mapping Φ^*).

Theorem 1.4.6. *[12, Theorem 4.4] Let* $\bar{u}$ *be a local minimizer of (1.19)-(1.20) satisfying Assumption 1.4.5. Then the optimality mapping* Φ^* *is strongly metrically subregular at* $(y_{\bar{u}}, \lambda_{\bar{u}})$ *in the metrics* $\mathcal{M}^*$ *and* $\mathcal{Z}^*$.

Some other complementary results were also proved in [12, Section 4], but those were more of a routine task. Theorem 1.4.6 truly represents a novelty in the optimal control context. It is also worth-mentioning that the elliptic equation studied in [12] was non-monotone and non-coercive, see [4, 5].

Bibliography

[1] Georgi Angelov, Alberto Domínguez Corella, and Vladimir M. Veliov. On the accuracy of the model predictive control method. *SIAM Journal on Control and Optimization*, 60(4):2469–2487, 2022.

[2] Jean-Pierre Aubin and Arrigo Cellina. *Differential inclusions*, volume 264 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1984. Set-valued maps and viability theory.

[3] Eduardo Casas. Second order analysis for bang-bang control problems of PDEs. *SIAM J. Control Optim.*, 50(4):2355–2372, 2012.

[4] Eduardo Casas, Mariano Mateos, and Arnd Rösch. Analysis of control problems of nonmontone semilinear elliptic equations. *ESAIM Control Optim. Calc. Var.*, 26:Paper No. 80, 21, 2020.

[5] Eduardo Casas, Mariano Mateos, and Arnd Rösch. Numerical approximation of control problems of non-monotone and non-coercive semilinear elliptic equations. *Numer. Math.*, 149(2):305–340, 2021.

[6] Eduardo Casas and Fredi Tröltzsch. On optimal control problems with controls appearing nonlinearly in an elliptic state equation. *SIAM J. Control Optim.*, 58(4):1961–1983, 2020.

[7] Eduardo Casas, Daniel Wachsmuth, and Gerd Wachsmuth. Sufficient second-order conditions for bang-bang control problems. *SIAM J. Control Optim.*, 55(5):3066–3090, 2017.

[8] Eduardo Casas, Daniel Wachsmuth, and Gerd Wachsmuth. Second-order analysis and numerical approximation for bang-bang bilinear control problems. *SIAM J. Control Optim.*, 56(6):4203–4227, 2018.

[9] R. Cibulka, A. L. Dontchev, M. I. Krastanov, and V. M. Veliov. Metrically regular differential generalized equations. *SIAM J. Control Optim.*, 56(1):316–342, 2018.

[10] R. Cibulka, A. L. Dontchev, and A. Y. Kruger. Strong metric subregularity of mappings in variational analysis and optimization. *J. Math. Anal. Appl.*, 457(2):1247–1282, 2018.

[11] Klaus Deckelnick and Michael Hinze. A note on the approximation of elliptic control problems with bang-bang controls. *Comput. Optim. Appl.*, 51(2):931–939, 2012.

[12] Alberto Domínguez Corella, Nicolai Jork, and Eduardo Casas. New assumptions for stability analysis in elliptic optimal control problems. available as a preprint in arxiv: https://doi.org/10.48550/arxiv.2205.03813.

[13] Alberto Domínguez Corella, Nicolai Jork, and Vladimir M. Veliov. Stability in affine optimal control problems constrained by semilinear elliptic partial differential equations. available as a preprint in arxiv: https://doi.org/10.48550/arxiv.2204.12964.

[14] Alberto Domínguez Corella, Marc Quincampoix, and Vladimir M. Veliov. Strong bi-metric regularity in affine optimal control problems. *Pure Appl. Funct. Anal.*, 6(6):1119–1137, 2021.

[15] A. L. Dontchev, M. I. Krastanov, and V. M. Veliov. On the existence of Lipschitz continuous optimal feedback control. *Vietnam J. Math.*, 47(3):579–597, 2019.

[16] Asen L. Dontchev. *Lectures on variational analysis*, volume 205 of *Applied Mathematical Sciences*. Springer, Cham, [2021] ©2021.

[17] Asen L. Dontchev and William W. Hager. Lipschitzian stability in nonlinear control and optimization. *SIAM J. Control Optim.*, 31(3):569–603, 1993.

[18] Asen L. Dontchev and R. Tyrrell Rockafellar. *Implicit functions and solution mappings*. Springer Monographs in Mathematics. Springer, Dordrecht, 2009. A view from variational analysis.

[19] Roland Griesse. Lipschitz stability of solutions to some state-constrained elliptic optimal control problems. *Z. Anal. Anwend.*, 25(4):435–455, 2006.

[20] William W. Hager. Multiplier methods for nonlinear optimal control. *SIAM J. Numer. Anal.*, 27(4):1061–1080, 1990.

[21] M. Hinze and C. Meyer. Stability of semilinear elliptic optimal control problems with pointwise state constraints. *Comput. Optim. Appl.*, 52(1):87–114, 2012.

[22] A. D. Ioffe. Metric regularity—a survey Part 1. Theory. *J. Aust. Math. Soc.*, 101(2):188–243, 2016.

[23] Bui Trong Kien, Nguyen Quoc Tuan, Ching-Feng Wen, and Jen-Chih Yao. L^∞-stability of a parametric optimal control problem governed by semilinear elliptic equations. *Appl. Math. Optim.*, 84(1):849–876, 2021.

[24] E. B. Lee and L. Markus. Foundations of optimal control theory. The SIAM Series in Applied Mathematics. New York-London-Sydney: John Wiley and Sons, Inc. xii, 576 p. (1967)., 1967.

[25] K. Malanowski and F. Tröltzsch. Lipschitz stability of solutions to parametric optimal control for elliptic equations. *Control Cybernet.*, 29(1):237–256, 2000.

[26] B. S. Mordukhovich and T. T. A. Nghia. Full Lipschitzian and Hölderian stability in optimization with applications to mathematical programming and optimal control. *SIAM J. Optim.*, 24(3):1344–1381, 2014.

[27] N. P. Osmolovskii and V. M. Veliov. Metric sub-regularity in optimal control of affine problems with free end state. *ESAIM Control Optim. Calc. Var.*, 26:Paper No. 47, 19, 2020.

[28] Jong-Shi Pang and David E. Stewart. Differential variational inequalities. *Math. Program.*, 113(2, Ser. A):345–424, 2008.

[29] Frank Pörner and Daniel Wachsmuth. An iterative Bregman regularization method for optimal control problems with inequality constraints. *Optimization*, 65(12):2195–2215, 2016.

[30] Frank Pörner and Daniel Wachsmuth. Tikhonov regularization of optimal control problems governed by semi-linear partial differential equations. *Math. Control Relat. Fields*, 8(1):315–335, 2018.

[31] Frank Pörner and Daniel Wachsmuth. Tikhonov regularization of optimal control problems governed by semi-linear partial differential equations. *Math. Control Relat. Fields*, 8(1):315–335, 2018.

[32] J. Preininger, T. Scarinci, and V. M. Veliov. Metric regularity properties in bang-bang type linear-quadratic optimal control problems. *Set-Valued Var. Anal.*, 27(2):381–404, 2019.

[33] Nguyen Thanh Qui and Daniel Wachsmuth. Stability for bang-bang control problems of partial differential equations. *Optimization*, 67(12):2157–2177, 2018.

[34] Nguyen Thanh Qui and Daniel Wachsmuth. Full stability for a class of control problems of semilinear elliptic partial differential equations. *SIAM J. Control Optim.*, 57(4):3021–3045, 2019.

[35] M. Quincampoix and V. M. Veliov. Metric regularity and stability of optimal control problems for linear systems. *SIAM J. Control Optim.*, 51(5):4118–4137, 2013.

[36] Marc Quincampoix, Teresa Scarinci, and Vladimir M. Veliov. On the metric regularity of affine optimal control problems. *J. Convex Anal.*, 27(2):511–535, 2020.

[37] Stephen M. Robinson. Strongly regular generalized equations. *Math. Oper. Res.*, 5(1):43–62, 1980.

[38] Daniel Veldman and Enrique Zuazua. Local stability and convergence of unconstrained model predictive control. *arXiv preprint arXiv:2206.01097*, 2022.

Chapter 2

Strong bi-Metric Regularity in Affine Optimal Control Problems

This chapter consists of the paper:

- Domínguez Corella, Alberto and Quincampoix, Marc and Veliov, Vladimir M.: Strong bi-metric regularity in affine optimal control problems.

It was published in Pure Applied Functional Analysis (ISSN 2189-3756): Volume 6, pages 1119-1137, year 2021. The author of this thesis contributed with the proof of the so-called uniform Euler discretization method in terms of sufficient second-order assumptions; he wrote the last section of the paper, and helped with the technical details of all other sections.

Strong bi-metric regularity in affine optimal control problems*

Alberto D. Corella† Marc Quincampoix‡ Vladimir Veliov§

Abstract

The paper presents new sufficient conditions for the property of strong bi-metric regularity of the optimality map associated with an optimal control problem which is affine with respect to the control variable (*affine problem*). The optimality map represents the system of first order optimality conditions (Pontryagin principle), and its regularity is of key importance for the qualitative and numerical analysis of optimal control problems. The case of affine problems is especially challenging due to the typical discontinuity of the optimal control functions. A remarkable feature of the obtained sufficient conditions is that they do not require convexity of the objective functional. As an application, the result is used for proving uniform convergence of the Euler discretization method for a family of affine optimal control problems.

Keywords: optimal control, metric regularity, affine problems, Euler discretization
AMS Clasification: 49K40, 49M25, 49J53

1 Introduction

Regularity properties of the system of first order necessary optimality conditions for optimization problems play a key role in qualitative analysis and reliable numerical treatment of such problems (see e.g. the books [1, 8, 4, 7]). For optimal control problems, the investigation of regularity properties of the map associated with the Pontryagin maximum principle (called further *optimality map*) was first initiated in [3], which deals with problems that satisfy the so-called coercivity condition. The latter, however is never fulfilled for problems in which both dynamics and cost are affine with respect to the control (further called *affine problems*). Results about strong metric sub-regularity of the optimality map for affine problems were obtained in the recent papers [10, 9]. The property of *strong metric regularity* (see e.g. [4, Chapter 3]) of the optimality map proved to be important for convergence and error estimates of numerical methods (discretizations, gradient projection, Newton method, etc.). However, more suitable for affine problems is a specific extension

*This research is supported by the Austrian Science Foundation (FWF) under grant No P31400-N32, by the Air Force Office of Scientific Research under award number FA9550-18-1-0254 , by the "FMJH Program Gaspard Monge in optimization and operation research" (PGMO 2015-2832H, PGMO 2016-1570H and PGMO 2018-0047H), and from the support to this program from EDF and EDF-THALES-ORANGE-CRITEO.

†Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, alberto.corella@tuwien.ac.at

‡Univ Brest, UMR CNRS 6205, Laboratoire de Mathématiques de Bretagne Atlantique 6, Avenue Victor Le Gorgeu, 29200 Brest, France, Marc.Quincampoix@univ-brest.fr

§Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, vladimir.veliov@tuwien.ac.at

the strong metric regularity introduced in [12] under the name *Strong bi-Metric Regularity* (Sbi-MR). The present paper investigates this property for Lagrange-type affine optimal control problems of the form

$$\min\Big\{J(u) := \int_0^T [w(t,x(t)) + \langle s(t,x(t)), u(t)\rangle]\,\mathrm{d}t\Big\}, \tag{1}$$

subject to

$$\dot{x}(t) = a(t,x(t)) + B(t,x(t))u(t), \quad x(0) = x^0, \tag{2}$$

$$u(t) \in U, \quad t \in [0,T]. \tag{3}$$

Here the state vector $x(t)$ belongs to $\mathbb{R}^n$, the control function u has values $u(t)$ that belong to a given set U in $\mathbb{R}^m$ for almost every (a.e.) $t \in [0,T]$. Correspondingly, w is a scalar function on $[0,T]\times\mathbb{R}^n$, s is an m-dimensional vector function ($\langle\cdot,\cdot\rangle$ denotes the scalar product), a and B are vector-/matrix-valued functions with appropriate dimensions. The initial state x^0 and the final time $T > 0$ are fixed. The set of feasible control functions u, denoted in the sequel by $\mathcal{U}$, consists of all Lebesgue measurable and bounded functions $u : [0,T] \to U$. Accordingly, the state trajectories x, that are solutions of (2) for feasible controls, are Lipschitz continuous functions on $[0,T]$.

It is well known that the Pontryagin (local) maximum principle can be written in the form of a generalized equation

$$0 \in F(y), \tag{4}$$

where $y = (x(\cdot), u(\cdot), p(\cdot))$ encapsulates the state function $x(\cdot)$, the control function $u(\cdot) \in \mathcal{U}$, and the adjoint (co-state) function $p(\cdot)$, and the inclusion $0 \in F(y)$ represents the state equation, the co-state equation, and the maximization condition in the maximum principle (the last being the inclusion of the derivative of the associated Hamiltonian with respect to the control in the normal cone to $\mathcal{U}$ at $u(\cdot)$). The detailed definition of the mapping F in (4), called further *optimality map* is given in the next section.

In the next paragraphs we remind the definition of Sbi-MR in the form used in [10] and [11]. Let (Y, d_Y), (Z, d_Z), $(\tilde{Z}, d_{\tilde{Z}})$ be metric spaces, with $\tilde{Z} \subset Z$ and $d_Z \le d_{\tilde{Z}}$ on $\tilde{Z}$.[1] Denote by $\mathbb{B}_Y(\hat{y}; a)$, $\mathbb{B}_Z(\hat{z}; b)$ and $\mathbb{B}_{\tilde{Z}}(\hat{z}; b)$ the closed balls in the metric spaces (Y, d_Y), (Z, d_Z) and $(\tilde{Z}, d_{\tilde{Z}})$ with radius $a > 0$ or $b > 0$ centered at $\hat{y}$ and $\hat{z}$, respectively.

Given a set-valued map $\Phi : Y \rightrightarrows Z$, $\operatorname{graph}\Phi := \{(y,z) \in Y \times Z : \ z \in \Phi(y)\}$ denotes the graph of Φ. The inverse map, $\Phi^{-1} : Z \rightrightarrows Y$, is the set-valued map defined as $\Phi^{-1}(z) := \{y \in Y : z \in \Phi(y)\}$.

Definition 1.1. *The set-valued map* $\Phi : Y \rightrightarrows Z$ *is* strongly bi-metrically regular *(Sbi-MR) (with disturbance space* $\tilde{Z}$*) at* $\hat{y} \in Y$ *for* $\hat{z} \in \tilde{Z}$ *with constants* $\kappa \ge 0$, $a > 0$ *and* $b > 0$, *if* $(\hat{y}, \hat{z}) \in \operatorname{graph}(\Phi)$ *and the following properties are fulfilled:*

(i) the map $\mathbb{B}_{\tilde{Z}}(\hat{z}; b) \ni z \mapsto \Phi^{-1}(z) \cap \mathbb{B}_Y(\hat{y}; a)$ *is single-valued;*

(ii) for all $z, z' \in \mathbb{B}_{\tilde{Z}}(\hat{z}; b)$

$$d_Y(\Phi^{-1}(z) \cap \mathbb{B}_Y(\hat{y}; a), \Phi^{-1}(z') \cap \mathbb{B}_Y(\hat{y}; a)) \le \kappa d_Z(z, z'). \tag{5}$$

We stress that the difference between this notion and the standard notion of strong metric regularity (see e.g. [4, Chapter 3]) is that the "disturbances" z have to belong to the smaller space, $\tilde{Z}$ (with the bigger norm), but the Lipschitz property in (ii) holds with respect to the smaller distance,

[1] This inequality can be understood as $d_Z(z) \le c\, d_{\tilde{Z}}(z)$ for every $z \in \tilde{Z}$, where c is a constant.

d_Z, in the right-side of (5). A detailed explanation of the reasons for the appropriateness of this definition is given in [11, Introduction].

Sufficient conditions for more specific problems and some applications of the Sbi-MR property are presented in [10] and [11]. The main aim of the present paper is to obtain new, more general, sufficient conditions for Strong bi-Metric Regularity (Sbi-MR) of the optimality map F in an appropriate space setting. A new feature of these conditions is that they involve not only the second derivative of the associated Hamiltonian with respect to the control, but also its first derivative. Thanks to that, they may be also fulfilled for problems with a non-convex objective functional, which is a new founding in the optimal control context, in general.

We present the sufficient conditions for Sbi-MR in Section 2 and give a detailed proof in Section 3. In Section 4 we specialize these conditions to the case of affine problems with bang-bang solutions and give an example where they apply to a non-convex problem. As an application, in Section 5 we prove that the obtained sufficient conditions imply uniform first order convergence of the Euler discretization scheme when applied to affine problems that are close enough to a reference one. This result is of importance, for example, for the justification of Model Predictive Control methods applied to affine problems.

2 Sufficient conditions for strong bi-metric regularity

We will use the following standard notations. The euclidean norm and the scalar product in $\mathbb{R}^n$ (the elements of which are regarded as column-vectors) are denoted by $|\cdot|$ and $\langle\cdot,\cdot\rangle$, respectively. The transpose of a matrix (or vector) E is denoted by $E^\top$. For a function $\psi : \mathbb{R}^p \to \mathbb{R}^r$ of the variable z we denote by $\psi_z(z)$ its derivative (Jacobian), represented by an $(r \times p)$-matrix. If $r = 1$, $\nabla_z \psi(z) = \psi_z(z)^\top$ denotes its gradient (a vector-column of dimension p). Also for $r = 1$, $\psi_{zz}(z)$ denotes the second derivative (Hessian), represented by a $(p \times p)$-matrix. For a function $\psi : \mathbb{R}^{p+q} \to \mathbb{R}$ of the variables (z, v), $\psi_{zv}(z, v)$ denotes its mixed second derivative, represented by a $(p \times q)$-matrix. The space $L^k([0,T],\mathbb{R}^r)$, with $k = 1, 2$ or $k = \infty$, consists of all (classes of equivalent) Lebesgue measurable r-dimensional vector-functions defined on the interval $[0,T]$, for which the standard norm $\|\cdot\|_k$ is finite. Often the specification $([0,T],\mathbb{R}^r)$ will be omitted in the notations. As usual, $W^{1,k} = W^{1,k}([0,T],\mathbb{R}^r)$ denotes the space of absolutely continuous functions $x : [0,T] \to \mathbb{R}^r$ for which the first derivative belongs to L^k. The norm in $W^{1,k}$ is defined as $\|x\|_{1,k} := \|x\|_k + \|\dot{x}\|_k$. Moreover, $I\!B_X(x;r)$ will denote the ball of radius r centered at x in a metric space X.

Allover the paper we use the abbreviation

$$f(t,x,u) = a(t,x) + B(t,x)u, \qquad g(t,x,u) = w(t,x) + \langle s(t,x), u\rangle. \tag{6}$$

For problem (1)–(3) we make the following assumption.

Assumption (A1). The set U is convex and compact; the functions $f : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ have the form as in (6) and are two times differentiable in (t,x), and the second derivatives are Lipschitz continuous.

Define the Hamiltonian associated with problem (1)–(3) as usual:

$$H(t,x,p,u) := g(t,x,u) + \langle p, f(t,x,u)\rangle, \quad p \in \mathbb{R}^n.$$

The local form of the Pontryagin maximum (here minimum) principle for problem (1)-(3) can be represented by the following optimality system for (x, u) and an absolutely continuous (here Lipschitz) function $p : [0, T] \to \mathbb{R}^n$: for a.e. $t \in [0, T]$

$$0 = -\dot{x}(t) + f(t, x(t), u(t)), \quad x(0) - x^0 = 0, \tag{7}$$

$$0 = \dot{p}(t) + \nabla_x H(t, x(t), p(t), u(t)), \quad p(T) = 0, \tag{8}$$

$$0 \in \nabla_u H(t, x(t), p(t), u(t)) + N_U(u(t)), \tag{9}$$

where the normal cone $N_U(u)$ to the set U at $u \in \mathbb{R}^m$ is defined in the usual way,

$$N_U(u) = \begin{cases} \{y \in \mathbb{R}^n \mid \langle y, v - u\rangle \le 0 \text{ for all } v \in U\} & \text{if } u \in U, \\ \emptyset & \text{otherwise.} \end{cases}$$

Assumption (A1) implies that there exists a number $M > 0$ such that for any $u \in \mathcal{U}$ the corresponding solution x of (7) and also the solution p of (8) exist on $[0, T]$ and

$$\max\{|x(t)|, |\dot{x}(t)|, |p(t)|, |\dot{p}(t)|\} \le M \quad \text{for a.e. } t \in [0, T]. \tag{10}$$

In what follows, $\bar{M}$ will be any number larger that M.

Let us introduce the metric spaces

$$Y := \{y = (x, p, u) \in W^{1,1} \times W^{1,1} \times L^1 \ : \ u \in \mathcal{U},\ x(0) = x^0,\ p(T) = 0, \text{ and } x, p, \dot{x}, \dot{p} \in I\!B_{L^\infty}(0; \bar{M})\}.$$

and

$$Z := L^\infty \times L^\infty \times L^\infty \text{ and } \quad \tilde{Z} := L^\infty \times L^\infty \times W^{1,\infty} \subset Z.$$

The distances in these spaces are induced by norms, therefore we keep the norm-notations: for $y = (x, p, u) \in Y$

$$\|y\| := \|x\|_{1,1} + \|p\|_{1,1} + \|u\|_1$$

and for $z = (\xi, \pi, \rho)$ in Z or in $\tilde{Z}$, respectively,

$$\|z\|_Z := \|\xi\|_1 + \|\pi\|_1 + \|\rho\|_\infty, \qquad \|z\|_\sim := \|\xi\|_\infty + \|\pi\|_\infty + \|\rho\|_{1,\infty}.$$

Notice that Y is a complete metric space, thanks to the compactness of the set U.

Now, we define the set-valued mapping $F : Y \rightrightarrows Z$ as

$$F(y) := \begin{pmatrix} -\dot{x} + f(\cdot, x, u) \\ \dot{p} + \nabla_x H(\cdot, y) \\ \nabla_u H(\cdot, y) \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ N_{\mathcal{U}}(u) \end{pmatrix}, \tag{11}$$

where $N_{\mathcal{U}}(u)$ is the normal cone to the set $\mathcal{U}$ of admissible controls at u, considered as a subset of L^∞:

$$N_{\mathcal{U}}(u) := \begin{cases} \emptyset & \text{if } u \notin \mathcal{U} \\ \{v \in L^\infty \ : \ v(t) \in N_U(u(t)) \text{ for a.e. } t \in [0, T]\} & \text{if } u \in \mathcal{U}. \end{cases}$$

Notice that $F(Y) \subset Z$, and $\nabla_u H(\cdot, y) \in \tilde{Z}$ thanks to the affine structure of the problem, namely, the independence of $\nabla_u H(\cdot, y)$ of u.

With these definitions, the necessary optimality conditions (7)–(9) take the form

$$F(y) \ni 0, \tag{12}$$

therefore F is called *optimality map* associated with problem (1)–(3). The main result in this paper is a sufficient condition for Sbi-MR of the optimality mapping $F : Y \rightrightarrows Z$ with perturbation space $\tilde{Z}$. To do this we fix a reference solution $\hat{y} = (\hat{x}, \hat{p}, \hat{u})$. We mention that such always exists since on assumption (A1) problem (1)–(3) has a solution. To shorten the notations we skip arguments with "hat" in functions, shifting the "hat" on the top of the notation of the function, so that $\hat{f}(t) := f(t, \hat{x}(t), \hat{u}(t))$, $\hat{s}(t) = s(t, \hat{x}(t))$, $\hat{H}(t) := H(t, \hat{x}(t), \hat{u}(t), \hat{p}(t))$, $\hat{H}(t, u) := H(t, \hat{x}(t), u, \hat{p}(t))$, etc. Moreover, denote

$$\hat{A}(t) := f_x(t, \hat{x}(t), \hat{u}(t)), \quad \hat{B}(t) := f_u(t, \hat{x}(t), \hat{u}(t)) = B(t, \hat{x}(t)), \quad \hat{\sigma}(t) := \nabla_u \hat{H}(t) = \hat{B}(t)^\top \hat{p}(t) + \hat{s}(t).$$

Let us introduce the following functional of $L^1 \ni \delta u \mapsto \Gamma(\delta u) \in \mathbb{R}$:

$$\Gamma(\delta u) := \int_0^T \Big[\langle \hat{H}_{xx}(t)\delta x(t), \delta x(t)\rangle + 2\langle \hat{H}_{ux}(t)\delta x(t), \delta u(t)\rangle \Big] \,\mathrm{d}t, \tag{13}$$

where δx is the solution of the equation $\dot{\delta x} = \hat{A}\delta x + \hat{B}\delta u$ with initial condition $\delta x(0) = 0$.

Assumption (A2). There exist numbers c_0, $\alpha_0 > 0$ and $\gamma_0 > 0$ such that

$$\int_0^T \langle \sigma(t), \delta u(t)\rangle \,\mathrm{d}t + \Gamma(\delta u) \geq c_0 \|\delta u\|_1^2,$$

for every $\delta u = u' - u$ with $u', u \in \mathcal{U} \cap B_{L^1}(\hat{u}; \alpha_0)$, and for every function $\sigma \in \mathbb{B}_{W^{1,\infty}}(\hat{\sigma}; \gamma_0) \cap (-N_{\mathcal{U}}(u))$.

Assumption (A2) will be analyzed and discussed in details in Section 4. Now we formulate the main theorem.

Theorem 2.1. *Let Assumption (A1) be fulfilled for problem (1)–(3) and let $\hat{y} = (\hat{x}, \hat{p}, \hat{u})$ be a solution of the optimality system (12) (with F defined in (11)) for which Assumption (A2) is fulfilled. Let, in addition, the matrix $\hat{H}_{ux}(t)\hat{B}(t)$ be symmetric for a.e. $t \in [0, T]$. Then the optimality map $F : Y \rightrightarrows Z$ is strongly bi-metrically regular at $\hat{y}$ for zero with disturbance space $\tilde{Z} \subset Z$.*

3 Proof of the main result

The proof of Theorem 2.1 consists of several steps.

Step 1. The following result (adapted to the present problem formulation, assumptions, and notations) was proved in [11, Theorem 3.1].[2]

[2] A Mayer problem is considered in [11], but the result also applies to Lagrange problems after a standard transformation. Moreover, the assumptions in [11] are somewhat weaker than (A1).

Theorem 3.1. *Let the assumptions in Theorem 2.1 be satisfied. Then strong bi-metric regularity of the set-valued map $y \mapsto F(y)$ at $\hat{y}$ for 0 (in the spaces as in Theorem 2.1) is equivalent to the strong bi-metric regularity of the map $y \mapsto L(y)$, at $\hat{y}$ for 0, where*

$$L(y) = \begin{pmatrix} -\dot{x} + \hat{f} + \hat{A}(x - \hat{x}) + \hat{B}(u - \hat{u}) \\ \dot{p} + \nabla_x \hat{H} + \hat{H}_{xy}(y - \hat{y}) \\ \nabla_u \hat{H} + \hat{H}_{uy}(y - \hat{y}) + N_{\mathcal{U}}(u) \end{pmatrix}.$$

The map L represents the partial linearization of F around $\hat{y} = (\hat{x}, \hat{p}, \hat{u})$. Thanks to the identity $\hat{H}_{uu} = 0$, L maps Y to Z, and moreover, $\hat{y}$ solves the inclusion $L(\hat{y}) \ni 0$.

To shorten the notations, we set for this section (skipping the dependence on t)

$$W := \hat{H}_{xx}, \quad S := \hat{H}_{ux}, \quad A := \hat{A} = \hat{f}_x, \quad B = \hat{B} = B(\hat{x}).$$

We remind the already introduced notation $\hat{\sigma} = \nabla_u \hat{H}$. Then, also having in mind the identity $\hat{H}_{uu} = 0$, we can recast the definition of $L(y)$ as

$$L(y) = \begin{pmatrix} -\dot{x} + \dot{\hat{x}} + A(x - \hat{x}) + B(u - \hat{u}) \\ \dot{p} - \dot{\hat{p}} + W(x - \hat{x}) + S^\top(u - \hat{u}) + A^\top(p - \hat{p}) \\ \hat{\sigma} + S(x - \hat{x}) + B^\top(p - \hat{p}) + N_{\mathcal{U}}(u) \end{pmatrix}.$$

Due to Assumption (A1), we have that $\dot{\hat{x}}$, A, $\dot{\hat{p}}$, W, $\hat{\sigma} \in L^\infty$, and B, $S \in W^{1,\infty}$. We remind that according to (12) and (11), $\hat{u}$ satisfies the inclusion $\hat{\sigma} + N_{\mathcal{U}}(\hat{u}) \ni 0$.

Step 2. Define the map $\Lambda : L^1 \times \tilde{Z} \to L^\infty$ in the following way: for $u \in L^1$ and $z = (\xi, \pi, \rho) \in \tilde{Z}$,

$$\Lambda(u, z) := \hat{\sigma} + S(x[u, z] - \hat{x}) + B^\top(p[u, z] - \hat{p}) - \rho, \tag{14}$$

where $(x[u, z], p[u, z])$ is the solution of the system

$$\dot{x} = \dot{\hat{x}} + A(x - \hat{x}) + B(u - \hat{u}) - \xi, \quad x(0) = x^0, \tag{15}$$
$$-\dot{p} = -\dot{\hat{p}} + W(x - \hat{x}) + S^\top(u - \hat{u}) + A^\top(p - \hat{p}) - \pi, \quad p(T) = 0. \tag{16}$$

Further we skip the argument z if $z = 0$, so that $x[u] = x[u, 0]$, $p[u] = p[u, 0]$, $\Lambda(u) := \Lambda(u, 0)$.

Lemma 3.2. *Strong bi-metric regularity of the set-valued map L at $\hat{y}$ for 0 (in the spaces as in Theorem 2.1) is equivalent to strong bi-metric regularity of the map $\Lambda(\cdot, 0) + N_{\mathcal{U}}(\cdot) : \mathcal{U} \rightrightarrows L^\infty$ at $\hat{u}$ for zero, with disturbance space $W^{1,\infty} \subset L^\infty$.*

Proof. We shall prove that the bi-metric regularity of the map $\Lambda + N_{\mathcal{U}}$ implies that of L, which will actually be used later. The proof of the converse is similar and simpler.

For any $z = (\xi, \pi, \rho) \in \tilde{Z}$ and $u \in L^\infty$ we have from (15) that

$$x[u, z] - x[u, 0] = l^x(\xi), \quad \|l^x(\xi)\|_{1,\infty} \le c_1 \|\xi\|_\infty, \quad \|l^x(\xi)\|_{1,1} \le c_1' \|z\|_Z, \tag{17}$$

where l^x is a linear map from L^∞ to $W^{1,\infty}$ and c_1 and c_1' are independent of u and z. Using this and (16), we obtain (also in a standard way) that

$$p[u, z] - p[u, 0] = l^p(\xi, \pi), \quad \|l^p(\xi, \pi)\|_{1,\infty} \le c_2(\|\xi\|_\infty + \|\pi\|_\infty), \quad \|l^p(\xi, \pi)\|_{1,1} \le c_2' \|z\|_Z, \tag{18}$$

where l^p is a linear map from $L^\infty \times L^\infty$ to $W^{1,\infty}$, and c_2 and c_2' are constants such as c_1. Notice that the second inequalities in (17) and (18) imply that

$$\max\{|x[u,z](t)|,\, |\dot{x}[u,z](t)|,\, |p[u,z](t)|,\, |\dot{p}[u,z](t)|\} \le M + c''(\|\xi\|_\infty + \|\pi\|_\infty) \quad \text{for a.e. } t \in [0,T],$$

where c'' is a constant. This will be used later to ensure that the appearing triples $(u, x[u,z], p[u,z])$ belong to the space Y.

We may represent

$$\Lambda(u,z) = \Lambda(u) + Q(z),$$

where

$$Q(z) = Sl^x(\xi) + B^\top l^p(\xi,\pi) - \rho, \quad \|Q(z)\|_{1,\infty} \le c_3\|z\|_\sim$$

is a linear map and c_3 is a constant.

The inclusion $L(y) \ni z$ can be equivalently reformulated as

$$x = x[u,z], \quad p = p[u,z], \quad \Lambda(u,z) + N_\mathcal{U}(u) \ni 0. \tag{19}$$

In view of the obtained representations, the last relations are equivalent to

$$x = x[u] + l^x(\xi), \quad p = p[u] + l^p(\xi,\eta), \quad \Lambda(u) + Q(z) + N_\mathcal{U}(u) \ni 0.$$

Having in mind the estimations for $\|l^x(\xi)\|_{1,\infty}$, $\|l^p(\xi,\pi)\|_{1,\infty}$ and $\|Q(z)\|_{1,\infty}$, obtaining Sbi-MR of L from that of $\Lambda + N_\mathcal{U}$ becomes a routine task. We will sketch the rest of the proof for completeness.

First we observe that there is a constant c_4 such that $\|Q(z)\|_\infty \le c_4\|z\|_Z$. Let κ, α and β be the constants in the definition of the Sbi-MR of the map $\Lambda + N_\mathcal{U}$. Fix

$$\bar\alpha = (c_1' + c_2')\bar\beta + \alpha, \qquad \bar\beta = \min\left\{\frac{\beta}{c_3}, \frac{\bar M - M}{c''}\right\}, \qquad \bar\kappa = c_1' + c_2' + c_4\kappa.$$

For any $z \in \tilde Z$ with $\|z\|_\sim \le \bar\beta$ we have $\|Q(z)\|_{1,\infty} \le \beta$. Then there exists a unique solution $u(z) \in I\!B_{L^1}(\hat u; \alpha)$ of the inclusion $\Lambda(u,z) + N_\mathcal{U}(u) \ni 0$. Moreover, for z_1, $z_2 \in \tilde Z$ with $\|z_i\|_\sim \le \bar\beta$ we have

$$\|u(z_1) - u(z_2)\|_1 \le \kappa\|Q(z_1 - z_2)\|_\infty \le c_4\kappa\|z_1 - z_2\|_Z.$$

From the first two relations in (19) we have for $x(z_i) = x[u(z_i), z_i]$ and $p(z_i) = p[u(z_i), z_i]$

$$\|x(z_1) - x(z_2)\|_{1,1} + \|p(z_1) - p(z_2)\|_{1,1} \le c_1'\|z_1 - z_2\|_Z + c_2'\|z_1 - z_2\|_Z.$$

Thus L is Sbi-MR at $\hat y$ for zero with constants $\bar\kappa$, $\bar\alpha$, $\bar\beta$. □

Step 3. According to Lemma 3.2, it is enough to prove Sbi-MR of $\Lambda + N_\mathcal{U}$ in the spaces specified in the formulation of the lemma. It is convenient to use the notation

$$\langle v, u\rangle_1 := \int_0^T \langle v(t), u(t)\rangle \,\mathrm{d}t$$

for the duality pairing of L^1 and L^∞, where $v \in L^\infty$ and $u \in L^1$. The map $\Lambda : L^1 \to L^\infty$ is linear and continuous, and we shall show that its derivative, Λ', satisfies the equality

$$\langle \Lambda'\delta u, \delta u\rangle_1 = \Gamma(\delta u), \quad \forall\, \delta u \in L^1, \tag{20}$$

where the mapping $\Gamma : L^1 \to \mathbb{R}$ is defined in (13).[3] In the notations introduced in this section the definition of Γ reads as

$$\Gamma(\delta u) = \langle W\delta x, \delta x\rangle_1 + 2\langle S\delta x, \delta u\rangle_1, \tag{21}$$

where δx is the solution of $\dot{\delta x} = A\delta x + B\delta u$ with $\delta x(0) = 0$. Let δp be the solution of the equation

$$-\dot{\delta p} = A^\top \delta p + W\delta x + S^\top \delta u, \quad \delta p(T) = 0.$$

Since $u \mapsto \Lambda(u) := \hat{\sigma} + S(x[u] - \hat{x}) + B^\top(p[u] - \hat{p})$ is linear, we deduce

$$\Lambda'(u)\delta u = S\delta x + B^\top \delta p. \tag{22}$$

Integrating by parts the expression $\langle \delta p, \dot{\delta x}\rangle_1$ we obtain the equality

$$\langle \delta p, A\delta x + B\delta u\rangle_1 = \langle \delta p, \dot{\delta x}\rangle_1 = -\langle \delta x, \dot{\delta p}\rangle_1 = \langle \delta x, A^\top \delta p + W\delta x + S^\top \delta u\rangle_1.$$

Hence,

$$\langle \delta p, B\delta u\rangle_1 = \langle \delta x, W\delta x + S^\top \delta u\rangle_1,$$

$$\langle B^\top \delta p, \delta u\rangle_1 = \langle W\delta x, \delta x\rangle_1 + \langle S\delta x, \delta u\rangle_1,$$

$$\langle S\delta x, \delta u\rangle_1 + \langle B^\top \delta p, \delta u\rangle_1 = \langle W\delta x, \delta x\rangle_1 + 2\langle S\delta x, \delta u\rangle_1 = \Gamma(\delta u),$$

which implies (20) in view of (22).

Equality (20) allows to reformulate the inequality in Assumption (A2) as

$$\int_0^T \langle \sigma(t), \delta u(t)\rangle \,\mathrm{d}t + \langle \Lambda'\delta u, \delta u\rangle_1 \geq c_0 \|\delta u\|_1^2 \tag{23}$$

with σ and δu as in (A2).

Step 4. Next, we will prove that for every $\alpha \in (0, \alpha_0)$ (see Assumption (A2)) and for every $\Delta \in W^{1,\infty}$ with $\|\Delta\|_{1,\infty} < c_0\alpha$ the inclusion

$$\Lambda(u) + N_{\mathcal{U}}(u) \ni \Delta \tag{24}$$

has a solution $\tilde{u} \in L^1$ satisfying $\|\tilde{u} - \hat{u}\|_1 < \alpha$. For this, we consider the inclusion

$$\Lambda(u) + N_{\mathcal{U} \cap \mathbb{B}_{L^1}(\hat{u};\alpha)}(u) \ni \Delta. \tag{25}$$

This inclusion represents the standard necessary optimality condition for the problem

$$\min \left\{ J_0(u) := \int_0^T \left[\frac{1}{2}\langle W(t)x[u](t), x[u](t)\rangle + \langle S(t)x[u](t), u(t)\rangle + \langle \Delta(t), u(t)\rangle\right] \mathrm{d}t \right\},$$

where $x[u]$ is defined around (15), with the control constraints $u \in \mathcal{U}$ and $u \in \mathbb{B}_{L^1}(\hat{u};\alpha)$. This is due to the well-known fact that $\Lambda(u)$ is the derivative of J_0 at u in L^1 (the proof of this fact uses a similar argument as the proof of the relation (20)). Due to the weak compactness of $\mathcal{U} \cap \mathbb{B}_{L^1}(\hat{u};\alpha)$ in L^1, this problem has a solution $\tilde{u}$, which then is a solution of (25).

[3] Similar representations are known, see e.g. in [6], but in the space L^2. Here the space setting is different and the specificity of the affine problem is essential.

Now we use the relation

$$N_{\mathcal{U}\cap \boldsymbol{B}_{L^1}(\hat{u};\alpha)}(u) = N_{\mathcal{U}}(u) + N_{\boldsymbol{B}_{L^1}(\hat{u};\alpha)}(u). \tag{26}$$

It follows from [2, Theorem 3.1], which, formulated for the particular space setting and sets, $\mathcal{U} \subset L^1$ and $\mathcal{V} := I\!B_{L^1}(\hat{u};\alpha) \subset L^1$, reads as follows: the equality (26) holds, provided that $\text{Epi}\, s_{\mathcal{U}} + \text{Epi}\, s_{\mathcal{V}}$ is weak* closed, where $\text{Epi}\, s_{\mathcal{W}}$ is the epigraph of $s_{\mathcal{W}}$ and $s_{\mathcal{W}} : L^\infty \to \mathbb{R}$ is the support function to the set $\mathcal{W} \subset L^1$, that is, $s_{\mathcal{W}}(l) := \sup_{w\in\mathcal{W}} \langle l, w\rangle_1$. Thus we need the following lemma.

Lemma 3.3. *For the sets $\mathcal{U}$ and $\mathcal{V}$ defined in the last paragraph, the set* $\text{Epi}\, s_{\mathcal{U}} + \text{Epi}\, s_{\mathcal{V}} \subset L^\infty \times \mathbb{R}$ *is weak* closed.*

Proof of the lemma. Let $(\xi_k, \lambda_k) \in \text{Epi}\, s_{\mathcal{U}}$ and $(\eta_k, \mu_k) \in \text{Epi}\, s_{\mathcal{V}}$ be arbitrary sequences such that $\xi_k + \eta_k \to \psi$ in the weak* topology and $\lambda_k + \mu_k \to \nu$. We will prove that $(\psi, \nu) \in \text{Epi}\, s_{\mathcal{U}} + \text{Epi}\, s_{\mathcal{V}}$.

We successively obtain the inequalities

$$\begin{aligned}
&\langle \eta_k, v\rangle_1 \le \mu_k, \qquad \text{for every } v \in \mathcal{V},\\
&\alpha\langle \eta_k, v\rangle_1 \le -\langle \eta_k, \hat{u}\rangle_1 + \mu_k \qquad \text{for every } v \in I\!B_{L^1}(0;1),\\
&\alpha\|\eta_k\|_\infty \le \lambda_k + \mu_k - \langle \xi_k + \eta_k, \hat{u}\rangle_1 - \lambda_k + \langle \xi_k, \hat{u}\rangle\\
&\qquad\qquad \le \lambda_k + \mu_k - \langle \xi_k + \eta_k, \hat{u}\rangle,
\end{aligned}$$

where in the last inequality we use that $\hat{u} \in \mathcal{U}$, thus $\langle \xi_k, \hat{u}\rangle \le \lambda_k$. Passing to the limit we obtain that $\alpha\|\eta_k\|_\infty \le \nu - \langle \psi, \hat{u}\rangle + 1$ for every sufficiently large k. Since the balls in L^∞ are weak* compact, we obtain that (a subsequence of) $\{\eta_k\}$ is weak* convergent to some $\eta \in L^\infty$. Then $\{\xi_k\}$ is also weak* convergent to some ξ.

Since $0 \in \mathcal{V}$, we have $\mu_k \ge s_{\mathcal{V}}(\eta_k) \ge \langle \eta_k, 0\rangle_1 = 0$. From $\hat{u} \in \mathcal{U}$, we get $\lambda_k \ge \langle \xi_k, \hat{u}\rangle_1 \ge \langle \xi, \hat{u}\rangle_1 - 1$ for k large enough. Because the sequence $\lambda_k + \mu_k$ is convergent, and hence bounded, we deduce that sequences λ_k and μ_k are bounded. So they converge respectively (up to a subsequence) to some real numbers λ and μ.

Passing to the limit in the inequality $\langle \xi_k, u\rangle \le \lambda_k$, which holds for every $u \in \mathcal{U}$, we obtain that $(\xi, \lambda) \in \text{Epi}\, s_{\mathcal{U}}$. Similarly, $(\eta, \mu) \in \text{Epi}\, s_{\mathcal{V}}$, which completes the proof of the lemma. □

Due to (26) and (25), there exists $\nu \in N_{\boldsymbol{B}_{L^1}(\hat{u};\alpha)}(\tilde{u})$ such that

$$\nu + \Lambda(\tilde{u}) - \Delta \in -N_{\mathcal{U}}(\tilde{u}),$$

hence,

$$\langle \nu, \hat{u} - \tilde{u}\rangle_1 + \langle \Lambda(\tilde{u}) - \Delta, \hat{u} - \tilde{u}\rangle_1 \ge 0.$$

We have $\langle \nu, \hat{u} - \tilde{u}\rangle_1 \le 0$ since $\hat{u} \in I\!B_{L^1}(\hat{u};\alpha)$. Thus

$$\langle \Lambda(\tilde{u}), \hat{u} - \tilde{u}\rangle_1 - \langle \Delta, \hat{u} - \tilde{u}\rangle_1 \ge 0.$$

Since Λ is linear and satisfies (20), and since $\Lambda(\hat{u}) = \hat{\sigma}$ in view of (14), we obtain that

$$\begin{aligned}
0 &\ge \langle \Lambda(\hat{u}), \tilde{u} - \hat{u}\rangle_1 + \langle \Lambda(\tilde{u}) - \Lambda(\hat{u}), \tilde{u} - \hat{u}\rangle_1 + \langle \Delta, \hat{u} - \tilde{u}\rangle_1\\
&= \langle \Lambda(\hat{u}), \tilde{u} - \hat{u}\rangle_1 + \langle \Lambda'(\tilde{u} - \hat{u}), \tilde{u} - \hat{u}\rangle_1 + \langle \Delta, \hat{u} - \tilde{u}\rangle_1\\
&= \langle \hat{\sigma}, \tilde{u} - \hat{u}\rangle_1 + \Gamma(\tilde{u} - \hat{u}) + \langle \Delta, \hat{u} - \tilde{u}\rangle_1.
\end{aligned}$$

Moreover, we have $\hat\sigma \in -N_{\mathcal U}(\hat u)$. Then Assumption (A2) in the form of (23) applied for $\delta u = \tilde u - \hat u$ and $\sigma = \hat\sigma$ implies that

$$0 \ge c_0 \|\tilde u - \hat u\|_1^2 + \langle \Delta, \hat u - \tilde u\rangle_1.$$

Hence,

$$\|\tilde u - \hat u\|_1 \le \frac{\|\Delta\|_\infty}{c_0} < \alpha.$$

Since $\tilde u$ belongs to the interior of $I\!\!B_{L^1}(\hat u;\alpha)$, thus $N_{B_{L^1}(\hat u;\alpha)}(\tilde u) = \{0\}$, we obtain that $\nu = 0$, therefore $\tilde u$ is a solution of the inclusion (24).

Step 5. First, we shall estimate $\|\Lambda(u_1) - \Lambda(u_2)\|_{1,\infty}$ for two functions $u_1, u_2 \in L^1$. Denote $\delta u = u_1 - u_2$, $\delta x = x[u_1] - x[u_2]$, $\delta p = p[u_1] - p[u_2]$. Then there is a constant c_1 independent of u_1 and u_2 such that

$$\|\delta x\|_\infty \le c_1 \|\delta u\|_1, \qquad \|\delta p\|_\infty \le c_1 \|\delta u\|_1.$$

Using the definition of Λ and Assumption (A1) we can estimate

$$\|\Lambda(u_1) - \Lambda(u_2)\|_\infty \le c_2 \|\delta u\|_1$$

with some constant c_2. Then

$$\begin{aligned} \Big\|\frac{\mathrm d}{\mathrm dt}(\Lambda(u_1) - \Lambda(u_2))\Big\|_\infty &\le \|\dot S \delta x + \dot B^\top \delta p\|_\infty + \|S(A\delta x + B\delta u) - B^\top(W\delta x + S^\top \delta u + A^\top \delta p)\|_\infty \\ &\le c_3 \|\delta u\|_1, \end{aligned}$$

where c_3 is another constant and in the last estimate we use the assumed symmetry of $SB = \hat H_{ux}\hat B$. Thus

$$\|\Lambda(u_1) - \Lambda(u_2)\|_{1,\infty} \le (c_2 + c_3)\|u_1 - u_2\|_1 =: c_4 \|u_1 - u_2\|_1. \tag{27}$$

Now we choose the number α in such a way that

$$0 < \alpha \le \alpha_0, \qquad c_0\alpha \le \alpha_0, \qquad (c_0 + c_4)\alpha \le \gamma_0.$$

Consider two disturbances $\Delta_1, \Delta_2 \in W^{1,\infty}$ with $\|\Delta_i\|_{1,\infty} < c_0\alpha$, and two solutions $u_1, u_2 \in I\!\!B_{L^1}(\hat u;\alpha)$ of (24) corresponding to Δ_1 and Δ_2, respectively. Similarly as in Step 4 we obtain the following chain of inequalities:

$$\begin{aligned} 0 &\ge \langle \Lambda(u_1) - \Delta_1, u_1 - u_2\rangle_1 \\ &= \langle \Lambda(u_2) - \Delta_2, u_1 - u_2\rangle_1 + \langle \Lambda(u_1) - \Lambda(u_2), u_1 - u_2\rangle_1 + \langle \Delta_2 - \Delta_1, u_1 - u_2\rangle_1 \\ &= \langle \sigma, u_1 - u_2\rangle_1 + \langle \Lambda'(u_1 - u_2), u_1 - u_2\rangle_1 + \langle \Delta_2 - \Delta_1, u_1 - u_2\rangle_1 \qquad (28) \\ &= \langle \sigma, u_1 - u_2\rangle_1 + \Gamma(u_1 - u_2) + \langle \Delta_2 - \Delta_1, u_1 - u_2\rangle_1, \qquad (29) \end{aligned}$$

where $\sigma := \Lambda(u_2) - \Delta_2$. Using (27) we obtain

$$\|\sigma - \hat\sigma\|_{1,\infty} \le \|\Lambda(u_2) - \Lambda(\hat u)\|_{1,\infty} + \|\Delta_2\|_{1,\infty} \le c_4\|u_2 - \hat u\|_1 + c_0\alpha < c_4\alpha + c_0\alpha \le \gamma_0. \tag{30}$$

Moreover, we have $\sigma = \Lambda(u_2) - \Delta_2 \in -N_{\mathcal U}(u_2)$ because u_2 solves the variational inequality (24) with $\Delta = \Delta_2$. Having in mind also that $\|u_2 - \hat u\|_1 < \alpha \le \alpha_0$, we can apply Assumption (A2) (in the form as in (23)) to the last line of (29). We obtain

$$0 \ge c_0 \|u_1 - u_2\|_1^2 + \langle \Delta_2 - \Delta_1, u_1 - u_2\rangle_1,$$

which implies that $\|u_1 - u_2\|_1 \le \frac{1}{c_0}\|\Delta_1 - \Delta_2\|_\infty$. This proves the Sbi-MR property of $\Lambda + N_{\mathcal U}$ with constants $\kappa = (c_0)^{-1}$, α, and $\beta = c_0\alpha$. The proof of Theorem 2.1 is complete.

4 Some special cases

We begin with few comments. Assumption (A2) with the particular choice $\sigma = \hat{\sigma}$, reads as

$$\int_0^T \langle \hat{\sigma}(t), \delta u(t) \rangle \, \mathrm{d}t + \Gamma(\delta u) \geq c_0 \|\delta u\|_1^2. \tag{31}$$

This inequality, required for all $\delta u \in \mathcal{U} - \hat{u}$, is shown in [9] to be sufficient for the property of *strong metric sub-regularity*, which is substantially weaker than Sbi-MR. Moreover, the condition[4]

$$\int_0^T \langle \hat{\sigma}(t), \delta u(t) \rangle \, \mathrm{d}t + \frac{1}{2}\Gamma(\delta u, \delta x) \geq c_0 \|\delta u\|_1^2, \quad \forall\, \delta u \in \mathcal{U} - \hat{u}, \quad \|\delta u\|_1 \text{ – sufficiently small},$$

is sufficient for strict local optimality of $\hat{u}$ in an L^1-neighborhood. This last condition is weaker than (31), as shown in [9].

Assumption (A2) is fulfillled on the following (more compact) one.

Assumption (A2'). There exist numbers c_0, $\alpha_0 > 0$ and $\gamma_0 > 0$ such that

$$\int_0^T |\langle \sigma(t), \delta u(t) \rangle| \, \mathrm{d}t + \Gamma(\delta u) \geq c_0 \|\delta u\|_1^2, \tag{32}$$

for every function $\sigma \in I\!B_{W^{1,\infty}}(\hat{\sigma}; \gamma_0)$ and for every $\delta u \in \mathcal{U} - \mathcal{U}$ with $\|\delta u\|_1 \leq \alpha_0$.

The implication (A2') $\Longrightarrow$ (A2) is obvious.

Now we focus on the first-order term in (32) under an additional condition introduced in [5] in a somewhat stronger form and for box-like sets U.

Assumption (B). The set U is a convex and compact polyhedron. Moreover, there exist numbers $\kappa > 0$ and $\tau > 0$ such that for every unit vector e parallel to some edge of U and for every $s \in [0, T]$ for which $\langle \hat{\sigma}(s), e \rangle = 0$ it holds that

$$|\langle \hat{\sigma}(t), e \rangle| \geq \kappa |t - s| \qquad t \in [s - \tau, s + \tau] \cap [0, T].$$

The next lemma claims that Assumption (B) remains valid for all functions σ close enough to $\hat{\sigma}$ in $W^{1,\infty}$.

Lemma 4.1. *Let assumptions (A1) and (B) be fulfilled. Then there exist numbers $\kappa' > 0$, $\tau' > 0$ and $\gamma' > 0$ such that for every function $\sigma \in I\!B_{W^{1,\infty}}(\hat{\sigma}; \gamma')$, for every unit vector e parallel to some edge of U and for every $s \in [0, T]$ for which $\langle \sigma(s), e \rangle = 0$, it holds that*

$$|\langle \sigma(t), e \rangle| \geq \kappa |t - s| \qquad t \in [s - \tau, s + \tau] \cap [0, T].$$

[4] The left-hand side in the next inequality is just the second order Taylor expansion of the objective functional $J(u)$ in (1).

Proof. The proof combines arguments from the proof of Proposition 3.4 in [11] and the proof of Proposition 3.1 in [9], therefore we only sketch it focusing on the differences with the proofs mentioned above.

First of all, Assumption (B) implies that the reference control $\hat{u}$ is piece-wise constant. This follows from the fact that $\langle \hat{\sigma}(t), e \rangle$ has not more than $T/\tau + 1$ zeros in $[0,T]$ and U has a finite number of edges. More details are given in the proof of Proposition 3.1 in [9].

From the definition of $\hat{\sigma}$, (A1) and the fact that $\hat{u}$ is a piece-wise constant function we obtain that $\hat{\sigma}$ has a piece-wise continuous derivative. Let us fix e as in Assumption (B), and denote $\hat{\sigma}_e := \langle \hat{\sigma}(t), e \rangle$. Let $\hat{s}_1, \ldots, \hat{s}_k$ be the zeros of $\hat{\sigma}_e$ in $[0,T]$. For $\delta > 0$ define

$$\Omega(\delta) := \cup_{i=1}^{k} [\hat{s}_i - \delta, \hat{s}_i + \delta].$$

Choose $\delta > 0$ so small that $\delta < \tau$ and there are no other points of discontinuity of $\dot{\hat{\sigma}}$ in $\Omega(\delta)$ except possibly $\hat{s}_1, \ldots, \hat{s}_k$. Denote

$$\dot{\hat{\sigma}}_e^-(\hat{s}_i) := \lim_{t \to \hat{s}_i - 0} \dot{\hat{\sigma}}(t), \quad \dot{\hat{\sigma}}_e^+(\hat{s}_i) := \lim_{t \to \hat{s}_i + 0} \dot{\hat{\sigma}}(t), \quad i = 1, \ldots, k.$$

By choosing $\delta > 0$ smaller, if needed, we may ensure that

$$|\dot{\hat{\sigma}}(t) - \dot{\hat{\sigma}}_e^-(\hat{s}_i)| \le \frac{\kappa}{4} \quad \text{for} \quad t \in [\hat{s}_i - \delta, \hat{s}_i], \qquad |\dot{\hat{\sigma}}(t) - \dot{\hat{\sigma}}_e^+(\hat{s}_i)| \le \frac{\kappa}{4} \quad \text{for} \quad t \in [\hat{s}_i, \hat{s}_i + \delta].$$

Then for every i and $t \in [\hat{s}_i - \delta, \hat{s}_i]$ we have from Assumption (B) that

$$\begin{aligned} \kappa|t - \hat{s}_i| &\le |\hat{\sigma}_e(t) - \hat{\sigma}_e(\hat{s}_i)| = \Big| \int_{\hat{s}_i}^{t} \dot{\hat{\sigma}}_e(\theta)\, \mathrm{d}\theta \Big| \le \int_{\hat{s}_i}^{t} |\dot{\hat{\sigma}}_e^-(\hat{s}_i)|\, \mathrm{d}\theta + \int_{\hat{s}_i}^{t} |\dot{\hat{\sigma}}_e^-(\hat{s}_i) - \dot{\hat{\sigma}}_e(\theta)|\, \mathrm{d}\theta \\ &\le |t - \hat{s}_i|\, |\dot{\hat{\sigma}}_e^-(\hat{s}_i)| + \frac{\kappa}{4}|t - \hat{s}_i| \end{aligned}$$

Hence,

$$|\dot{\hat{\sigma}}_e^-(\hat{s}_i)| \ge \frac{3\kappa}{4}.$$

Analogously we obtain the same estimate for $|\dot{\hat{\sigma}}_e^+(\hat{s}_i)|$.

Obviously there exists $\eta > 0$ such that $|\hat{\sigma}_e(t)| \ge \eta$ for every $t \in [0,T] \setminus \Omega(\delta/2)$. By choosing the number $\gamma \in (0, \kappa/4]$ sufficiently small we have that for every $\sigma \in I\!B_{W^{1,\infty}}(\hat{\sigma}; \gamma)$ the function $\sigma_e = \langle \sigma, e \rangle$ has no zeros in $[0,T] \setminus \Omega(\delta/2)$. Now let us take an arbitrary σ as in the last sentence. Let s be an arbitrary zero of σ_e in $[0,T]$. Then there exists $\hat{s}_i$ such that $|s - \hat{s}_i| \le \delta/2$. For $t \in [s - \delta/2, s + \delta/2]$ we can estimate

$$|\sigma_e(t)| = \Big| \int_s^t \dot{\sigma}_e(\theta)\, \mathrm{d}\theta \Big| \ge \Big| \int_s^t \dot{\hat{\sigma}}_e(\theta)\, \mathrm{d}\theta \Big| - \int_s^t |\dot{\sigma}_e(\theta) - \dot{\hat{\sigma}}_e(\theta)|\, \mathrm{d}\theta \Big| \ge \Big| \int_s^t \dot{\hat{\sigma}}_e(\theta)\, \mathrm{d}\theta \Big| - \gamma|t - s|.$$

For the last integral we have

$$\Big| \int_s^t \dot{\hat{\sigma}}_e(\theta)\, \mathrm{d}\theta \Big| \ge \Big| \int_s^t \zeta(\theta)\, \mathrm{d}\theta \Big| - \int_s^t |\dot{\hat{\sigma}}_e(\theta) - \zeta(\theta)|\, \mathrm{d}\theta,$$

where $\zeta(\theta)$ is either $\dot{\hat{\sigma}}_e^-(\hat{s}_i)$ or $\dot{\hat{\sigma}}_e^+(\hat{s}_i)$ depending on whether $\theta < \hat{s}_i$ or $\theta > \hat{s}_i$. Thus we can estimate

$$|\sigma_e(t)| \ge \frac{3\kappa}{4}|t - s| - \frac{\kappa}{4}|t - s| - \gamma|t - s| \ge \frac{\kappa}{4}|t - s|.$$

Thus we obtain the claim of the lemma with $\kappa' = \kappa/4$, $\tau' = \delta/2$ and $\gamma' = \gamma$. □

Proposition 4.2. *Let assumptions (A1) and (B) be fulfilled. Then there exist numbers* c_0, $\alpha_0 > 0$ *and* $\gamma_0 > 0$ *such that*

$$\int_0^T |\langle \sigma(t), \delta u(t) \rangle| \,\mathrm{d}t \geq c_0 \|\delta u\|_1^2, \tag{33}$$

for every function $\sigma \in \mathbb{B}_{W^{1,\infty}}(\hat{\sigma}; \gamma_0)$ *and for every* $\delta u \in \mathcal{U} - \mathcal{U}$ *with* $\|\delta u\|_1 \leq \alpha_0$.

Having at hand Lemma 4.1, the proof repeats that of Proposition 3.1 in [9].

Remark 4.3. A more slightly precise modification of the proof of Lemma 4.1 shows that the number κ' can be taken as any number smaller than κ (from Assumption (B)). Moreover, the constant c_0 in Proposition 4.2 is directly related with number κ' (thus with κ). In the simplest case of scalar control and $U = [u_1, u_2]$ is straightforward. As obtained in the proof of Lemma 4.1, Assumption (B) implies in this case that $\hat{\sigma}$ has finite number of zeros, $\hat{s}_1, \ldots, \hat{s}_k$, and $\dot{\hat{\sigma}}$ is piece-wise continuous. If the number Q satisfies

$$\liminf_{t \to \hat{s}_i} |\dot{\hat{\sigma}}(t)| \geq Q \quad 1 = 1, \ldots, k,$$

(the liminf is taken over all t at which the derivative exists) then a simple calculation shows that the claim of Proposition 4.2 holds with any number $c_0 \leq Q/(8k(u_2 - u_1))$.

Example 1. This example shows that Sbi-MR of the optimality mapping may hold even for problems that are non-convex, namely, the objective functional J in (1) is even directionally non-convex at the optimal control $\hat{u}$. Consider the problem

$$\min \Big\{ J(u) := \int_0^1 \Big[-\frac{\alpha}{2}(x(t))^2 - \beta x(t) + u(t) \Big] \,\mathrm{d}t \Big\},$$

subject to

$$\dot{x} = u, \quad x(0) = 0, \quad u(t) \in [0, 1].$$

Here α and β are positive parameters satisfying $\beta > 1$, $2\alpha \leq \beta$.

The solution of the adjoint equation $\dot{p} = \alpha x + \beta$, $p(1) = 0$ is strictly monotone increasing and the switching function, $\sigma(t) = p(t) + 1$, is positive at $t = 1$. This implies that only optimal control has the structure

$$\hat{u}(t) = \begin{cases} 1 & \text{for } t \in [0, \tau], \\ 0 & \text{for } t \in (\tau, 1]. \end{cases}$$

The corresponding solutions of the primal and the adjoint equations are

$$\hat{x}(t) = \begin{cases} t & \text{for } t \in [0, \tau], \\ \tau & \text{for } t \in (\tau, 1], \end{cases} \quad \text{and} \quad \hat{p}(t) = \begin{cases} \frac{\alpha}{2}(\tau^2 + t^2) + \beta t - \alpha\tau - \beta & \text{for } t \in [0, \tau], \\ t(\alpha\tau + \beta) - \alpha\tau - \beta & \text{for } t \in (\tau, 1]. \end{cases}$$

A simple calculation shows that for $\beta > 1$ the optimal control $\hat{u}$ has

$$\tau = \frac{-(\beta - \alpha) + \sqrt{(\beta - \alpha)^2 + 4\alpha(\beta - 1)}}{2\alpha} \in (0, 1).$$

For the corresponding switching function $\hat{\sigma} = \hat{p} + 1$ we have $\dot{\hat{\sigma}}(\tau) = \alpha\tau + \beta > \beta$. Then Assumption (B) is fulfilled with $\kappa < \beta$. According to Remark 4.3, we have

$$\int_0^1 |\hat{\sigma}(t)\delta u(t)|\,\mathrm{d}t \geq \frac{\beta}{2}\|\delta u\|_1^2 \qquad \forall\, \delta u \in \mathcal{U} - \mathcal{U} \text{ with a sufficiently small } \|\delta u\|_1.$$

Moreover,

$$\Gamma(\delta u) = -\int_0^1 \alpha(\delta x(t))^2\,\mathrm{d}t = -\alpha\int_0^1 \Big(\int_0^t \delta u(s)\,\mathrm{d}s\Big)^2\,\mathrm{d}t \geq -\alpha\|\delta u\|_1^2.$$

Thus for $2\alpha < \beta$ Assumption (A2') is fulfilled and the optimality mapping for the considered problem is Sbi-MR at $(\hat{x}, \hat{u}, \hat{p})$ for zero. On the other hand, considering again the expression for the second variation Γ, we see that $\Gamma(\delta u) < 0$, except some specially constructed control variations δu. Thus the objective functional $J(u)$ in this example is not convex even directionally at the solution point $\hat{u}$.

5 An application: uniform convergence of the Euler discretization

In this section we prove that the sufficient conditions for Sbi-MR (this is essentially Assumption (A2)) imply a property that can be called *uniform strong sub-regularity* concerning a family of optimal control problems "neighboring" a given reference problem. This property is shown to imply a *uniform* error estimate for the accuracy of the Euler discretization scheme, applied to any of the problems of the family.

We consider again the reference problem (1)-(3) together with the fixed solution $(\hat{x}, \hat{p}, \hat{u})$ of its optimality system (7)–(9). Assumptions (A1) and (A2) will hold, with the additional supposition in (A1) that f and g are time-invariant.

Together with the reference problem, we consider a family of problems of the same kind, each defined by a pair of time-invariant functions $\pi := (\tilde{f}, \tilde{g})$ satisfying Assumption (A1) (with f and g replaced with $\tilde{f}$ and $\tilde{g}$). Any such pair will be called admissible, and $(\mathcal{P}_\pi)$ will denote the problem corresponding to the pair π, that is, the problem

$$\min_{u\in\mathcal{U}} \left\{ \int_0^T \tilde{g}(x(t), u(t))\,\mathrm{d}t \right\} \tag{34}$$

subject to

$$\dot{x}(t) = \tilde{f}(x(t), u(t)), \quad x(0) = x^0. \tag{35}$$

Due to relation (10), we restrict our consideration to admissible pairs π defined on the set $D := I\!\!B_{\mathbb{R}^n}(0, \bar{M}) \times U$. Given a positive number ρ, we denote by $\mathcal{H}_\rho$ the set of all admissible pairs $\pi = (\tilde{f}, \tilde{g})$ such that

$$\|\tilde{f} - f\|_{1,\infty} + \|\tilde{g} - g\|_{1,\infty} \leq \rho, \tag{36}$$

where the $W^{1,\infty}$-norms are taken for functions defined on the set D.

For a given $\pi = (\tilde{f}, \tilde{g}) \in \mathcal{H}_\rho$, we consider the mapping $\Phi_\pi : Y \to Z$ defined by

$$\Phi_\pi(x, p, u) = \begin{pmatrix} \dot{x} - \tilde{f}(x, u) \\ \dot{p} + \nabla_x \tilde{H}(x, p, u) \\ \nabla_u \tilde{H}(x, p, u) + N_{\mathcal{U}}(u) \end{pmatrix} \tag{37}$$

where $\tilde{H}$ is the Hamiltonian corresponding to the pair π, and where as before $N_{\mathcal{U}}(u) \subset L^\infty$ is the normal cone to the set $\mathcal{U}$ of admissible controls at u. The following lemma is technical.

Lemma 5.1. *Let $\pi = (\tilde{f}, \tilde{g})$ belong to $\mathcal{H}_\rho$ and $\varphi_\pi : Y \to Z$ be defined as*

$$\varphi_\pi(x,p,u) = \begin{pmatrix} \varphi_\pi^1(x,p,u) \\ \varphi_\pi^2(x,p,u) \\ \varphi_\pi^3(x,p,u) \end{pmatrix} := \begin{pmatrix} f(x,u) - \tilde{f}(x,u) \\ \nabla_x \tilde{H}(x,p,u) - \nabla_x H(x,p,u) \\ \nabla_u \tilde{H}(x,p,u) - \nabla_u H(x,p,u) \end{pmatrix}. \tag{38}$$

There exists a positive constant c such that

$$d_Z(\varphi_\pi(y), 0) \le c\rho \quad \forall y \in Y. \tag{39}$$

Proof. Let $y = (x,p,u) \in Y$. We estimate each one of the components of $\varphi_\pi(y)$. First,

$$\|\varphi_\pi^1(y)\|_1 = \|f(x,u) - \tilde{f}(x,u)\|_1 \le T\rho.$$

In a similar way,

$$\|\varphi_\pi^2(y)\|_1 = \|\nabla_x \tilde{H}(x,p,u) - \nabla_x H(x,p,u)\|_1 \le \|\tilde{f}_x - f_x\|_1 \|p\|_\infty + \|\tilde{g}_x - g_x\|_1 \le (\bar{M}+1)T\rho.$$

Analogously,

$$\|\varphi_\pi^3(y)\|_\infty \le (\bar{M}+1)\rho.$$

The result follows. ☐

We remind the notion of Strong Metric sub-Regularity (SMsR) for a set-valued mapping $\Phi : Y \to Z$. We make use of this notion in the following results.

Definition 5.2. *A set valued mapping $\Phi : Y \to Z$ is Strong Metric sub-Regular (SMsR) at y^* for zero if $0 \in \Phi(y^*)$ and there exist $a, b > 0$ and $\kappa > 0$ such that for any $z \in B_Z(0,b)$ and any solution $y \in B_Y(y^*, a)$ of the inclusion $z \in \Phi(y)$ it holds that $d_Y(y, y^*) \le \kappa d_Z(z, 0)$. We call a, b and κ the parameters of SMsR.*

According to Theorem 3.1 in [9], Assumption (A2) implies that the optimality map F in (11) is SMsR at $\hat{y}$ for zero (see Section 4). We fix its parameters $a, b > 0$ and $\kappa > 0$ of SMsR.

Proposition 5.3. *Let π belong to $\mathcal{H}_\rho$ and $y^* \in B_Y(\hat{y}, a)$ be a solution of problem $(\mathcal{P}_\pi)$. There exists a positive constant κ' such that*

$$d_Y(\hat{y}, y^*) \le \kappa' \rho, \tag{40}$$

for all sufficiently small ρ.

Proof. We can write $\Phi_\pi = \varphi_\pi + F$, where φ_π is the map (38) in Lemma 5.1 and F is the optimality mapping (11). Let $c > 0$ be the constant in that lemma, so that $d_Z(\varphi_\pi(y), 0) \le c\rho$ for all $y \in Y$. We can choose ρ small enough to ensure $\varphi_\pi(y) \in B_Z(0,b)$ for all $y \in Y$. Since y^* is a solution of problem $(\mathcal{P}_\pi)$, the inclusion $0 \in \varphi_\pi(y^*) + F(y^*)$ is satisfied. By SMsR, we have the desired inequality with $\kappa' := c\kappa$. ☐

Analogously as we defined the functional Γ; given a $\pi \in \mathcal{H}$ and a reference solution y^* of problem $(\mathcal{P}_\pi)$, we consider the functional $\Gamma_\pi : L^1 \to \mathbb{R}$ defined in terms of π and y^* as in (13). Explicitly,

$$\Gamma_\pi(\delta u) = \int_0^T \Big[\langle \tilde{H}_{xx}(y^*(t))\delta x(t), \delta x(t)\rangle + 2\langle\, \tilde{H}_{ux}(y^*(t))\delta x(t), \delta u(t)\rangle\Big]\, \mathrm{d}t,$$

where δx is the solution of the equation $\dot{\delta x}(t) = \tilde{f}_x(x^*(t), u^*(t))\delta x(t) + \tilde{f}_u(x^*(t), u^*(t))\delta u(t)$ with initial condition $\delta x(0) = 0$.

The following lemma establishes an estimation involving the functionals Γ_π and Γ.

Lemma 5.4. *Let π belong to $\mathcal{H}_\rho$ and $y^* \in B_Y(\hat{y}, a)$ be a solution of problem $(\mathcal{P}_\pi)$. There exists a constant $\eta > 0$ such that*

$$|\Gamma(v - u^*) - \Gamma_\pi(v - u^*)| \le \eta\, \rho \|v - u^*\|_1^2 \quad \forall v \in \mathcal{U},$$

for all sufficiently small ρ.

Proof. Using Proposition 5.3 and the Lipschitz continuity of the functions involved, we can find positive constants c_w and c_s such that

$$\|\hat{H}_{xx} - \tilde{H}^*_{xx}\|_1 \le c_w \rho, \tag{41}$$

and

$$\|\hat{H}_{ux} - \tilde{H}^*_{ux}\|_\infty \le c_s \rho. \tag{42}$$

Let $v \in \mathcal{U}$ and $v' = v - u^*$, we denote by $\delta\hat{x}$ and δx^* the solutions of

$$\dot{x} = \hat{A}x + \hat{B}v', \;\; x(0) = 0, \qquad \dot{x} = \tilde{A}^* x + \tilde{B}^* v', \;\; x(0) = 0, \tag{43}$$

respectively. There exist positive constants d_1 and d_2 such that

$$\max\{\|\delta\hat{x}\|_\infty, \|\delta x^*\|_\infty\} \le d_1 \|v'\|_1, \tag{44}$$

and

$$\|\delta\hat{x} - \delta x^*\|_\infty \le d_2 \rho \|v'\|_1. \tag{45}$$

Now,

$$\begin{aligned}
|\Gamma(v') - \Gamma_\pi(v')| &\le \left| \int_0^T \Big[\langle\, \hat{H}_{xx}\delta\hat{x}, \delta\hat{x}\rangle - \langle\, \tilde{H}^*_{xx}\delta x^*, \delta x^*\rangle\Big] \right| + 2 \left| \int_0^T \Big[\langle\, \hat{H}_{ux}\delta\hat{x} - \tilde{H}^*_{ux}\delta x^*, v'\rangle\Big] \right| \\
&\le \int_0^T |\langle\, \hat{H}_{xx}\delta\hat{x}, \delta\hat{x} - \delta x^*\rangle| + \int_0^T |\langle\, \hat{H}_{xx}\delta\hat{x} - \tilde{H}^*_{xx}\delta x^*, \delta x^*\rangle| + 2\int_0^T |\langle\, \hat{H}_{ux}\delta\hat{x} - \tilde{H}^*_{ux}\delta x^*, v'\rangle| \\
&\le \|\hat{H}_{xx}\delta\hat{x}\|_1 \|\delta\hat{x} - \delta x^*\|_\infty + \big(\|\hat{H}_{xx}(\delta\hat{x} - \delta x^*)\|_1 + \|(\hat{H}_{xx} - \tilde{H}^*_{xx})\delta x^*\|_1\big)\|\delta x^*\|_\infty + \\
&\quad 2\big(\|\hat{H}_{ux}(\delta\hat{x} - \delta x^*)\|_\infty + \|(\hat{H}_{ux} - \tilde{H}^*_{ux})\delta x^*\|_\infty\big)\|v'\|_1.
\end{aligned}$$

Taking (41)-(45) into account, the result follows. □

Theorem 5.5. *There exist $\zeta, \tilde{a}, \tilde{b} > 0$ and $\tilde{\kappa} > 0$ such that if $\pi \in \mathcal{H}_\zeta$ and $y^* \in B_Y(\hat{y}, a)$ is a solution for problem $(\mathcal{P}_\pi)$, then the map Φ_π is SMsR at y^* for zero with parameters $\tilde{a}, \tilde{b}, \tilde{\kappa}$.*

Proof. Let c_0, $\alpha_0 > 0$ and $\gamma_0 > 0$ be the numbers in Assumption (A2). If y^* is a solution for problem $(\mathcal{P}_\pi)$, we have $0 \in \Phi_\pi(y^*)$. By Proposition 5.3, there exists $\zeta > 0$ such that for any $\pi \in \mathcal{H}_\zeta$, $d_Y(\hat{y}, y^*) < \kappa' \zeta$ for some constant $\kappa' > 0$. We consider ζ small enough to guarantee $\|\hat{\sigma} - \tilde{\sigma}^*\|_{1,\infty} \le \gamma_0$ and $\|\hat{u} - u^*\|_1 < \alpha_0$. Let $\tilde{\alpha}_0 := \alpha_0 - \|\hat{u} - u^*\|_1$, so $B_{L^1}(u^*; \tilde{\alpha}_0) \subset B_{L^1}(\hat{u}; \alpha_0)$.

By Assumption (A2),

$$\int_0^T \langle \tilde{\sigma}^*, v - u^* \rangle + \Gamma(v - u^*) \ge c_0 \|v - u^*\|_1^2 \quad \forall v \in \mathcal{U} \cap B_{L^1}(u^*; \tilde{\alpha}_0), \tag{46}$$

or

$$\int_0^T \langle \tilde{\sigma}^*, v - u^* \rangle + \Gamma_\pi(v - u^*) \ge c_0 \|v - u^*\|_1^2 + \Big[\Gamma_\pi(v - u^*) - \Gamma(v - u^*)\Big] \quad \forall v \in \mathcal{U} \cap B_{L^1}(u^*; \tilde{\alpha}_0).$$

Taking Lemma 5.4 into account, we can choose ζ smaller if needed to ensure

$$\Gamma_\pi(v - u^*) - \Gamma(v - u^*) \ge -\frac{c_0}{2} \|v - u^*\|_1^2 \quad \forall v \in \mathcal{U} \cap B_{L^1}(u^*; \tilde{\alpha}_0).$$

Thus,

$$\int_0^T \langle \sigma^*, v - u^* \rangle + \Gamma_\pi(v - u^*) \ge \frac{c_0}{2} \|v - u^*\|_1^2 \quad \forall v \in \mathcal{U} \cap B_{L^1}(u^*; \tilde{\alpha}_0). \tag{47}$$

Let L be a bound for the Lipschitz constants of f, g and H their first derivatives in x, and H_{xu}, H_{up}. It is easy to see that $\tilde{L} := L + 2(1 + \bar{M})\zeta$ is a bound for the Lipschitz constants of $\tilde{f}, \tilde{g}$ and $\tilde{H}$, their first derivatives in x, and $\tilde{H}_{xu}, \tilde{H}_{up}$ for all $\pi \in \mathcal{H}_\zeta$. Analogously, we can find a bound $\tilde{M}$, depending on ζ and $\bar{M}$, for the functions $\tilde{f}, \tilde{H}$ and their derivatives (see Remark 2.1 in [9]). Finally, (47) implies that the hypotheses of Theorem 3.1 in [9] are fulfilled. We conclude that Φ_π is SMsR at y^* for zero. According to that theorem, the parameters of SMsR can be chosen as depending only on $\tilde{M}, T, c_0$ and $\tilde{L}$. This completes the proof. □

From now on, we only consider elements $\pi \in \mathcal{H}_\zeta$, since the latter theorem ensures that each map Φ_π is SMsR at a solution $y^* \in B_Y(\hat{y}, a)$ for zero. This automatically ensures that y^* is the unique local solution in $B_Y(\hat{y}, a)$ of the inclusion $0 \in \Phi_\pi(y)$.

Let $\{t_n\}_{n=0}^N$ be a grid on $[0, T]$ with equally spaced nodes and a step size h, that is, $t_k = kT/N$ for $i = 0, \ldots, N$. Given a $\pi \in \mathcal{H}_\zeta$, the discrete time problem $(\mathcal{P}_\pi^h)$ obtained by the Euler discretization is

$$\min_{u \in U^N} \left[h \sum_{i=0}^{N-1} \tilde{g}(x_i, u_i) \right] \tag{48}$$

subject to

$$x_{i+1} = x_i + h \tilde{f}(x_i, u_i), \quad x_0 = x^0. \tag{49}$$

The local form of the discrete time minimum principle implies that for any locally optimal solution (x, u) of problem $(\mathcal{P}_\pi^h)$ there exists a vector $p = (p_0, \ldots, p_N)$ such that

$$x_{i+1} = x_i + h \tilde{f}(x_i, u_i), \; x_0 = x^0, \tag{50}$$

$$\lambda_i = \lambda_{i+1} + h \nabla_x \tilde{H}(x_i, u_i, p_{i+1}), \; p_N = 0, \tag{51}$$

$$0 \in \nabla_u \tilde{H}(x_i, u_i, p_{i+1}) + N_U(u_i), \tag{52}$$

where i runs between 0 and $N-1$. Let (x^h, u^h) be a solution of problem $(\mathcal{P}_\pi^h)$ and p^h the corresponding co-state vector, so that $y^h = (x^h, p^h, u^h)$ satisfies (50)-(52). In order to compare this solution to the reference solution of $y^* = (x^*, p^*, u^*)$ of the continuous-time problem $(\mathcal{P}_\pi)$, we embed the sequence (x^h, p^h, u^h) into the space $W^{1,1} \times W^{1,1} \times L^1$ considering $y_h = (x_h, p_h, u_h)$ defined by

$$x_h(t) := x_i^h + \frac{t-t_i}{h}(x_{i+1}^h - x_i^h), \quad u_h(t) := u_i^h, \quad p_h(t) := p_i^h + \frac{t-t_i}{h}(p_{i+1}^h - p_i^h), \tag{53}$$

for $t \in [t_i, t_{i+1})$, $i = 0, \dots, N-1$.

We need the following technical assumption to apply results in [9]. It is a crucial assumption, at least because it may happen that y_h is close to some other local solution of the continuous-time problem, and we have to eliminate this possibility.

Assumption (C1). Let $\pi \in \mathcal{H}_\zeta$. We assume that problem $(\mathcal{P}_\pi)$ has a solution y^* in $B_Y(\hat{y}, a)$. Moreover, the embedded solution y_h in (53) of problem $(\mathcal{P}_\pi^h)$ belongs to $B_Y(y^*, \tilde{a})$ for all sufficiently small h.

The following theorem is a direct consequence of Theorem 5.5 and Theorem 5.1 in [9].

Theorem 5.6. *There exists a positive constant C such that for all $\pi \in \mathcal{H}_\zeta$ for which Assumption (C1) holds, the estimate*

$$\|x_h - x^*\|_{1,1} + \|p_h - p^*\|_{1,1} + \|u_h - u^*\|_1 \le Ch \tag{54}$$

holds for all sufficiently small h.

Proof. By Theorem 5.5, the parameters $\tilde{a}, \tilde{b}, \tilde{\kappa}$ of SMsR of Φ_π at y^* for zero are the same for all $\pi \in \mathcal{H}_\zeta$ satisfying Assumption (C1).

Let $\pi \in \mathcal{H}_\zeta$. In order to make use of the SMsR property of the map Φ_π, we have to estimate the residuals

$$\begin{aligned}
&\Delta_1 := \dot{x}_h - \tilde{f}(x_h, u_h), \\
&\Delta_2 := \dot{p}_h + \nabla_x \tilde{H}(x_h, p_h, u_h), \\
&\Delta_3 := \nabla_u \tilde{H}(x_i^h, p_i^h, u_i^h) - \nabla_u \tilde{H}(x_h, p_h, u_h), \; t \in [t_i, t_{i+1}), \; i = 0, \dots, N-1.
\end{aligned}$$

Repeating the calculations in the proof of Theorem [9, Theorem 5.1], we obtain

$$\max\{\|\Delta_1\|_1, \|\Delta_2\|_1, \|\Delta_3\|_\infty\} \le \max\{1, T\}\, \tilde{L}(1 + 2\tilde{M})h, \tag{55}$$

where $\tilde{L}, \tilde{M}$ are the numbers in Theorem 5.5. We can choose $h_0 > 0$ depending on $\tilde{L}, \tilde{M}, T$ and b so that $\|\Delta_1\|_1 + \|\Delta_2\|_1 + \|\Delta_3\|_\infty \le b$ for all $h \le h_0$. The claim follows from the SMsR property of Φ_π with $C := 3\kappa(1 + 2\tilde{M})\tilde{L} \max\{1, T\}$. The proof is complete since this holds for any arbitrary $\pi \in \mathcal{H}_\zeta$ satisfying Assumption (C1). $\square$

References

[1] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000.

[2] R.S. Burachik and V. Jeyakumar. A simple closure condition for the normal cone intersection formula. *Proc. Amer. Math. Soc.*, 133(6):1741–1748, 2005.

[3] A.L. Dontchev and W.W. Hager. Lipschitzian stability in nonlinear control and optimization. *SIAM J. Control Optim.*, 31(3):569–603, 1993.

[4] A.L. Dontchev and R.T. Rockafellar. *Implicit functions and solution mappings*. Springer Monographs in Mathematics. Springer, Dordrecht, 2009. A view from variational analysis.

[5] U. Felgenhauer. On stability of bang-bang type controls. *SIAM J. Control Optim.*, 41(6):1843–1867, 2003.

[6] W.W. Hager. Multiplier methods for nonlinear optimal control. *SIAM J. Numer. Anal.*, 27(4):1061–1080, 1990.

[7] A.D. Ioffe. *Variational analysis of regular mappings*. Springer Monographs in Mathematics. Springer, Cham, 2017. Theory and applications.

[8] D. Klatte and B. Kummer. *Nonsmooth equations in optimization*, volume 60 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 2002. Regularity, calculus, methods and applications.

[9] N.P. Osmolovskii and V.M. Veliov. Metric sub-regularity in optimal control of affine problems with free end state. *ESAIM. Control, Optimisation and Calculus of Variations*. Forthcoming. DOI: https://doi.org/10.1051/cocv/2019046 . Available as Research Report 2019-04, ORCOS, TU Wien, 2019.

[10] J. Preininger, T. Scarinci, and V.M. Veliov. Metric regularity properties in bang-bang type linear-quadratic optimal control problems. *Set-Valued Var. Anal.*, 27(2):381–404, 2019.

[11] M. Quincampoix, T. Scarinci, and V.M. Veliov. On the metric regularity of affine optimal control problems. *Journal of Convex Analysis*, 27(2), 2020.

[12] M. Quincampoix and V.M. Veliov. Metric regularity and stability of optimal control problems for linear systems. *SIAM J. Control Optim.*, 51(5):4118–4137, 2013.

Chapter 3

On the Accuracy of the Model Predictive Control

This chapter consists of the paper:

- Angelov, Georgi and Domínguez Corella, Alberto and Veliov, Vladimir M.: On the accuracy of the model predictive control method.

It was published in SIAM Journal on Control and Optimization : Volume 4, pages 2469-2487, year 2022. The paper was written between the second and the third author in several stages, and in deep collaboration. The thesis author (second) was very involved in all sections of the paper as well as with the main result. The author wrote parts of each section, and helped the first author to write the section concerned with the numerical analysis, although he was not involved in the computations (creating and running the program).

On the accuracy of the model predictive control method*

Georgi Angelov† Alberto Domínguez Corella‡ Vladimir M. Veliov§

Abstract

The paper investigates the accuracy of the Model Predictive Control (MPC) method for finding on-line approximate optimal feedback control for Bolza type problems on a fixed finite horizon. The predictions for the dynamics, the state measurements, and the solution of the auxiliary open-loop control problems that appear at every step of the MPC method may be inaccurate. The main result provides an error estimate of the MPC-generated solution compared with the optimal open-loop solution of the "ideal" problem, where all predictions and measurements are exact. The technique of proving the estimate involves an extension of the notion of strong metric sub-regularity of set-valued maps and utilization of a specific new metric in the control space, which makes the proof non-standard. The result is specialized for two problem classes: coercive problems, and affine problems.

Keywords: optimal control, Lagrange problem, model predictive control, metric sub-regularity

MSC Classification: 93B45, 49M99, 49J40, 47J20

1 Introduction

Model Predictive Control (MPC) is a powerful method for approximate on-line feedback control, widely used in industrial applications and recently in digital engine control and microelectronics, see, e.g., [21, 26, 29]. On the other hand, the rigorous mathematical theory investigating the scope of validity and the efficiency of the MPC method under appropriate assumptions is still underdeveloped (see Remark 1.1 below for some related comments). The present paper contributes to this theory by investigating the accuracy of a version of the MPC method applied to a finite horizon optimal control problem (the so-called *economic MPC* with shrinking horizon).

To set the stage, we consider the following optimal control problem, further denoted by $\mathscr{P}_p(0,x_0)$:

$$\min_{u\in\mathscr{U}} \Big\{ J_p(u) := g_T(x(T)) + \int_0^T g(p(t),x(t),u(t))\,\mathrm{d}t \Big\}, \tag{1}$$

subject to

$$\dot{x}(t) = f(p(t),x(t),u(t)) \quad x(0) = x_0. \tag{2}$$

*This research is supported by the Austrian Science Foundation (FWF) under grant No I4571.

†Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, georgi.angelov@tuwien.ac.at

‡Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, alberto.corella@tuwien.ac.at

§Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, vladimir.veliov@tuwien.ac.at

Here the state vector $x(t)$ belongs to $\mathbb{R}^n$ and the control function $u(\cdot)$ belongs to the set $\mathscr{U}$ of all Lebesgue measurable functions $u : [0,T] \to U$, where $U \subset \mathbb{R}^m$. The function p represents an uncertain time-dependent parameter which is known to belong to a set Π of bounded Lebesgue measurable functions $p : [0,T] \to \mathbb{R}^l$. Correspondingly, f, g, and g_T are defined on $\mathbb{R}^l \times \mathbb{R}^n \times \mathbb{R}^m$ with values in $\mathbb{R}^n$ and $\mathbb{R}$, respectively. The initial state $x_0 \in \mathbb{R}^n$ and the final time $T > 0$ are fixed.

A version of the MPC method (called further MPC algorithm) applied to the above problem is described in detail in Subsection 3.1. Here, we briefly present the main result given in Theorem 3.1, Subsection 3.2. It is assumed that for some particular parameter function, $\hat{p}$, equation (2) represents a real system, therefore problem (1)–(2) with $p = \hat{p}$ (that is, problem $\mathscr{P}_{\hat{p}}(0,x_0)$) is called *reference problem*. However, $\hat{p}$ and the initial state x_0 are not assumed to be exactly known. The MPC algorithm generates a control function by solving a sequence of auxiliary open-loop optimal control problems. Given a mesh $0 = t_0 < t_1 < \ldots < t_N = T$, the auxiliary problem at the k-th step is of the same kind as (1)–(2), but on the shorter time-interval $[t_k, T]$. A prediction p for the parameter function on $[t_k, T]$ is given, and the initial state at t_k is obtained by measuring the real systems state at time t_k. Both the prediction and the measurement may be inaccurate. In addition, the auxiliary problem at the k-th step is only approximately solved, which is another source of error. The approximate optimal control in the k-th auxiliary problem is only applied to the ''real" system on the interval $[t_k, t_{k+1}]$, then the procedure is further repeated on the next interval, resulting at the end in what is called the MPC- generated control.

The main result in the paper gives an estimate of the difference between the MPC-generated control and the optimal open-loop control for the reference problem (the latter corresponding to the "ideal" scenario where the prediction, the measurement, and the solution of the auxiliary problems are all exact). A remarkable feature of the estimation is that the overall error of the MPC-generate control depends on the *average* of the errors appearing at the steps of the algorithm, thus occasional relatively large errors in the prediction or measurement do not substantially damage the MPC-generated control. Another interesting feature of the overall error is that for some classes of problems it depends linearly on the averaged errors appearing at the steps of the method, while for other classes, the estimate of the overall error depends on the square root of the averaged errors (and this estimate is sharp). We mention that an estimate of the difference between the MPC-generated control and the realization of the optimal *feedback* control in the reference problem is obtained [12]. In [11], this result is extended to a comparison with the optimal open-loop control in the reference problem, as in the present paper. However, in both quoted papers the results are obtained within a much more restrictive framework: a single prediction is used which does not change from step to step, the results only apply to the Euler discretization of the auxiliary problems, and most importantly, the reference optimal control problem is assumed *coercive* (see Subsection 4.1 for the notion). In fact, the main goal of the present paper is to extend the results about the accuracy of the MPC method beyond the coercive case, especially for affine control problems.

Remark 1.1. There is an amount of literature related to robustness of the MPC method, see e.g., [23, 19] and several chapters in [26]. Stabilization or target problems for discrete dynamics are usually considered for particular classes of problems, and various notions of robustness and approaches are investigated – see e.g. the overview paper [3, Chapter by Raković]. The result in the present paper, as explained above, is more related to the paper [20]. In that paper the accuracy of the MPC method is investigated (for a general class of discrete-time problems) in terms of the objective value, only. Of course, this is the primal accuracy criterion, however, it is of interest to know how close are the MPC-generated control and trajectory to the one optimal open-loop ones for the case of exact and complete information. The present paper contributes to this issue and, most importantly, employs and further develops the approach based on the general property of strong sub-regularity of a map associated with the problem into consideration.

The main result (the error estimate in Theorem 3.1) is obtained under general assumptions; the most demanding one is the requirement that the map associated with the first order optimality conditions (called *optimality map*) for the reference problem is *strongly metrically sub-regular* in an appropriate space setting. In Subsection 2.1, we extend the abstract notion of strong metric sub-regularity of a set-valued map by involving two metrics in the domain of the map. In Subsection 2.2, we define the optimality map and the space setting. In the control space (which is a projection of the domain of the optimality map) we introduce a specific new metric, which is of key importance for the analysis of the MPC method and may be useful in other contexts.

The proof of Theorem 3.1 (given in Appendix) is non-trivial and substantially differs from all proofs of error estimates in optimal control that the authors know.

Section 4 presents or recalls sufficient conditions for the extended strong metric sub-regularity of the optimality map for two non-intersecting classes of problems: for *coercive problems* and for *affine problems*. The paper concludes with an example where the MPC algorithm is applied to a spacecraft stabilization problem. The numerical results confirm the theoretical error estimate and its sharpness.

2 Preliminaries

2.1 Strong sub-regularity

Here we introduce a notion which extends the property of *Strong (metric) sub-Regularity* (Ss-R) (see, e.g., [16, Chapter 3.9] and the recent paper [5]). Namely, we consider a general metric space $\mathscr{Y}$ in which two metrics are defined: d and d^*, and another metric space $\mathscr{Z}$ with a metric $d_{\mathscr{Z}}$. We denote by $\boldsymbol{B}(y;\alpha)$ the closed ball with radius $\alpha \geq 0$ in $(\mathscr{Y}, d)$ centered at y, by $\boldsymbol{B}_*(y;\alpha)$ the ball with radius $\alpha \geq 0$ in $(\mathscr{Y}, d^*)$, and similarly, $\boldsymbol{B}_{\mathscr{Z}}(z;\alpha)$ is the respective ball in $\mathscr{Z}$.

Definition 2.1. *A set-valued map $\Phi : \mathscr{Y} \rightrightarrows \mathscr{Z}$ is called Ss-R at $(\hat{y}, \hat{z}) \in \mathscr{Y} \times \mathscr{Z}$ (with respect to the metrics d and d^*) if $\hat{z} \in \Phi(\hat{y})$ and there are positive constants α, β and κ (called further parameters of Ss-R) such that*

$$d^*(y,\hat{y}) \leq \kappa d_{\mathscr{Z}}(z,\hat{z}) \quad \text{for all } y \in \boldsymbol{B}(\hat{y};\alpha),\ z \in \Phi(y) \cap \boldsymbol{B}_{\mathscr{Z}}(\hat{z};\beta).$$

The Ss-R property plays a fundamental role in the error analysis of numerical methods. It was introduced under this name in [15], but has also been used under several other names (see also [22, Chapter 1] for the related but stronger property of strong upper regularity). A more detailed historical account can be found in [5, Section 1]. The extension with two metrics in $\mathscr{Y}$, presented above, is essential for the applications in the present paper.

The following simple claim is a modification of [5, Theorem 2.1] for the case of two metrics in Y.

Proposition 2.2. *Assume that $\mathscr{Z}$ is a linear space and $d_{\mathscr{Z}}$ is a shift-invariant metric in $\mathscr{Z}$. Assume, in addition, there exists a number $\gamma > 0$ such that $d(y_1,y_2) \leq \gamma d^*(y_1,y_2)$ for every $y_1, y_2 \in \mathscr{Y}$. Let $\Phi : \mathscr{Y} \rightrightarrows \mathscr{Z}$ be Ss-R at $(\hat{y},\hat{z})$ with parameters α', β' and κ'. Let the positive numbers $\varepsilon, \mu, \kappa, \alpha, \beta$ satisfy the relations*

$$\alpha \leq \alpha', \quad \beta + \mu\alpha + \varepsilon \leq \beta', \quad \mu\gamma\kappa' < 1, \quad \kappa = \frac{\kappa'}{1-\mu\gamma\kappa'}. \tag{3}$$

Then for every function $\varphi : \mathscr{Y} \to \mathscr{Z}$ that satisfies the conditions

$$d_{\mathscr{Z}}(\varphi(\hat{y}),0) \leq \varepsilon, \quad d_{\mathscr{Z}}(\varphi(y),\varphi(\hat{y})) \leq \mu\, d(y,\hat{y}) \quad \forall y \in \boldsymbol{B}(\hat{y};\alpha),$$

the map $\varphi + \Phi$ is strongly sub-regular at $(\hat{y}, \hat{z} + \varphi(\hat{y}))$ with parameters α, β and κ.

Proof. Obviously, $(\hat{y}, \hat{z} + \varphi(\hat{y})) \in \operatorname{graph}(\varphi + \Phi)$. Let us fix arbitrarily $z \in \boldsymbol{B}_{\mathscr{Z}}(\hat{z}; \beta)$ and $y \in \boldsymbol{B}(\hat{y}; \alpha) \subset \boldsymbol{B}(\hat{y}; \alpha')$ such that $z \in \varphi(y) + \Phi(y)$. Then $z - \varphi(y) \in \Phi(y)$, and

$$\begin{aligned} d_{\mathscr{Z}}(z - \varphi(y), \hat{z}) &\le d_{\mathscr{Z}}(z, \hat{z}) + d_{\mathscr{Z}}(\varphi(y), \varphi(\hat{y})) + d_{\mathscr{Z}}(\varphi(\hat{y}), 0) \\ &\le \beta + \mu\alpha + \varepsilon \le \beta'. \end{aligned}$$

Due to the Ss-R property of Φ we estimate

$$\begin{aligned} d^*(y, \hat{y}) &\le \kappa' d_{\mathscr{Z}}(z - \varphi(y), \hat{z}) \le \kappa' d_{\mathscr{Z}}(z, \hat{z} + \varphi(\hat{y})) + \kappa' d_{\mathscr{Z}}(\varphi(y), \varphi(\hat{y})) \\ &\le \kappa' d_{\mathscr{Z}}(z, \hat{z} + \varphi(\hat{y})) + \kappa' \mu\, d(y, \hat{y}) \le \kappa' d_{\mathscr{Z}}(z, \hat{z} + \varphi(\hat{y})) + \kappa' \mu \gamma d^*(y, \hat{y}), \end{aligned}$$

which implies the claim of the theorem due to the definition of κ in (3). □

2.2 The optimality map

Problem $\mathscr{P}_p(0, x_0)$ given by (1)–(2) will be considered under the following assumptions.

Assumption (A1). The set U is convex and compact. The functions f, g, and g_T are two times differentiable with respect to (x, u), these functions and their first and second derivatives in (x, u) are Lipschitz continuous (with respect to (p, x, u)) with a Lipschitz constant L.

For any $p \in \Pi$, along with problem $\mathscr{P}_p(0, x_0)$ we consider the family, denoted by $\mathscr{P}_p(\tau, x_\tau)$, consisting of problems which have the same form as (1)–(2) but with the initial time 0 replaced with any $\tau \in [0, T)$ and x_0 replaced with any $x_\tau \in \mathbb{R}^n$. Of course, then only the restriction of the parameter p to $[\tau, T]$ matters.

Assumption (A2). For every $u \in \mathscr{U}$, $x_0 \in \mathbb{R}^n$, and $p \in \Pi$ equation (2) has a solution x on $[0, T]$ (which is then unique due to Assumption (A1)). For every $\tau \in [0, T)$, $x_\tau \in \mathbb{R}^n$ and $p \in \Pi$ problem $\mathscr{P}_p(\tau, x_\tau)$ has an optimal solution.

Since the analysis in this paper is local, only local Lipschitz continuity of the functions mentioned in Assumption (A1) is needed; we assume global Lipschitz continuity to avoid routine technicalities. Assumption (A2) is also stronger than necessary, again for the sake of transparency. Local existence around a reference parameter and trajectory (to be introduced later) suffices.

Remark 2.3. Optimality in the last assumption means local optimality of the objective functional with respect to the L^1-norm of the controls. In fact, it is only needed that any solution (x, u) of $\mathscr{P}_p(\tau, x_\tau)$ satisfies, together with an absolutely continuous (*adjoint*) function $\lambda : [\tau, T] \to \mathbb{R}^n$, the *optimality (Pontryagin) system*

$$\begin{aligned} 0 &= -\dot{x}(t) + f(p(t), x(t), u(t)), \quad x(\tau) - x_\tau = 0, && (4) \\ 0 &= \dot{\lambda}(t) + \nabla_x H(p(t), x(t), \lambda(t), u(t)), \quad \lambda(T) = \nabla g_T(x(T)), && (5) \\ 0 &\in \nabla_u H(p(t), x(t), \lambda(t), u(t)) + N_U(u(t)), && (6) \end{aligned}$$

where

$$N_U(u) := \begin{cases} \{q \in \mathbb{R}^n \mid \langle q, v - u \rangle \le 0 \text{ for all } v \in U\} & \text{if } u \in U, \\ \emptyset & \text{if } u \notin U \end{cases}$$

is the normal cone to U at u, and the Hamiltonian H is defined as usual:

$$H(p,x,\lambda,u) := g(p,x,u) + \langle \lambda, f(p,x,u)\rangle.$$

Next, we reformulate the optimality system in functional spaces. The space $L^q(\tau,T)$, $q = 1,2,\dots,\infty$, of vector functions on $[\tau,T]$ (with any fixed dimension) has the usual meaning, with the norm denoted by $\|\cdot\|_q$. The space of all absolutely continuous vector functions on $[\tau,T]$ is denoted by $W^{1,1}(\tau,T)$, with the norm $\|x\|_{1,1} = \|x\|_1 + \|\dot{x}\|_1$. The notations of norms do not include the time horizon, but it will be clear from the context. For the same reason we often skip the time horizon from the notations of spaces.

Denote

$$Y_\tau := W^{1,1}(\tau,T)\times W^{1,1}(\tau,T)\times \mathscr{U}_\tau, \qquad Z_\tau := L^1(\tau,T)\times\mathbb{R}^n\times L^1(\tau,T)\times\mathbb{R}^n\times L^\infty(\tau,T),$$

where $\mathscr{U}_\tau = \{u \in L^\infty(\tau,T) : u(t)\in U \text{ for a.e. } t\in[\tau,T]\}$ is the set of admissible control functions on $[\tau,T]$, thus $\mathscr{U}_0 = \mathscr{U}$. We also set $Y := Y_0$ and $Z := Z_0$. The metrics in Y_τ and Z_τ are given in terms of norms as follows: for $y = (x,\lambda,u)\in Y_\tau$ and $z = (\xi,\nu,\eta,\pi,\rho)\in Z_\tau$

$$d(y,0) := \|y\| := \|x\|_{1,1} + \|\lambda\|_{1,1} + \|u\|_1, \qquad d_Z(z,0) := \|z\| := \|\xi\|_1 + |\nu| + \|\eta\|_1 + |\pi| + \|\rho\|_\infty.$$

In addition, we define in Y a second metric, d^*, as follows. Let $\Gamma\subset[0,T]$ be a fixed finite set. For $u_1,u_2\in\mathscr{U}_\tau$, denote

$$d^*(u_1,u_2) := \inf\{\varepsilon>0 : |u_1(t)-u_2(t)|\le\varepsilon \text{ for a.e. } t\in[0,T]\setminus(\Gamma+[-\varepsilon,\varepsilon])\}. \tag{7}$$

Somewhat overloading the notation, we define in Y the shift-invariant metric

$$d^*(y,0) := \|x\|_{1,1} + \|\lambda\|_{1,1} + d^*(u,0).$$

Lemma 2.4. *For every $u_1,u_2\in\mathscr{U}$ it holds that*

$$\|u_1-u_2\|_1 \le \gamma\, d^*(u_1,u_2),$$

where $\gamma := \max\{1, T+2M\,\mathrm{diam}\,(U)\}$ and M is the number of points in Γ.

The proof is straightforward. Since $\gamma\ge1$, we also have $\|y\|\le\gamma d^*_\tau(y,0)$ for any $y\in Y_\tau$.

Below we use the same notation for the Nemytskii operator and for its generating function: $f(p,x,u)(t) = f(p(t),x(t),u(t)$, $\nabla_x H(p,x,\lambda,u)(t) = \nabla_x H(p(t),x(t),\lambda(t),u(t))$, etc. For any $p\in\Pi$, $\tau\in[0,T)$, and $x_\tau\in\mathbb{R}^n$ define on Y_τ the set-valued map

$$\Phi_{(p,\tau,x_\tau)}(y) = F_{(p,\tau,x_\tau)}(y) + \begin{pmatrix} 0\\ 0\\ 0\\ 0\\ N_{\mathscr{U}_\tau}(u)\end{pmatrix}, \qquad F_{(p,\tau,x_\tau)}(y) = \begin{pmatrix} -\dot{x}+f(p,x,u)\\ x(\tau)-x_\tau\\ \dot{\lambda}+\nabla_x H(p,x,\lambda,u)\\ \lambda(T)-\nabla g(x(T))\\ \nabla_u H(p,x,\lambda,u)\end{pmatrix},$$

where now $N_{\mathscr{U}_\tau}(u)$ is the normal cone to $\mathscr{U}_\tau$ at u in the space $L^1(\tau,T)$, that is,

$$\begin{aligned} N_{\mathscr{U}_\tau}(u) &:= \begin{cases} \{l\in L^\infty(\tau,T) : \int_\tau^T \langle l(t), v(t)-u(t)\rangle\,\mathrm{d}t \le 0 \text{ for all } v\in\mathscr{U}_\tau\} & \text{if } u\in\mathscr{U}_\tau,\\ \emptyset & \text{if } u\notin\mathscr{U}_\tau\end{cases}\\ &= \{l\in L^\infty(\tau,T) : l(t)\in N_U(u(t)) \text{ for a.e. } t\in[\tau,T]\}.\end{aligned}$$

With these notations one can recast the optimality system for problem $\mathscr{P}_p(\tau, x_\tau)$ as

$$0 \in \Phi_{(p,\tau,x_\tau)}(x, \lambda, u),$$

therefore the map $\Phi_{(p,\tau,x_\tau)}$ is called *optimality map*. Obviously, due to the compactness of the set U, $\Phi_{(p,\tau,x_\tau)}$ is a set-valued map from Y_τ to Z_τ.

Let us fix a reference parameter $\hat{p} \in \Pi$ and denote by $(\hat{x}, \hat{u})$ a solution of problem $\mathscr{P}_{\hat{p}}(0, x_0)$ (see Remark 2.3). Let $\hat{\lambda}$ be the corresponding adjoint function, so that the triplet $\hat{y} := (\hat{x}, \hat{\lambda}, \hat{u})$ satisfies the optimality system (4)–(6) with $\tau = 0$ and $x_\tau = x_0$, or equivalently, the inclusion $0 \in \Phi_{(\hat{p},0,x_0)}(\hat{y})$. The following assumption plays a key role in the error analysis of the MPC method presented in the next section.

Assumption (A3). The map $\Phi_{(\hat{p},0,x_0)} : Y \rightrightarrows Z$ is strongly sub-regular at $(\hat{y}, 0)$ (in the metrics $\|\cdot\|$ and d^* in Y) with parameters $\hat{\alpha}, \hat{\beta}$ and $\hat{\kappa}$.

The finite set Γ that appears in the definition of the metric d^* is arbitrary, but will be appropriately specified in Section 4 for several classes of problems, together with sufficient conditions for (A3).

3 The accuracy of the model predictive control method

This section presents the main result: the estimate of the accuracy of the model predictive control method, beginning with the description of the method in the context of finite horizon optimal control.

3.1 The model predictive control method

The MPC, applied to optimal control problems containing uncertain parameters, is a method for approximation of an optimal feedback control in real time by successively solving open-loop optimal control problems. Each of these open-loop problems involve measurements of the current system state and predictions for the uncertain parameters. In the next three paragraphs we present a version of the MPC method.

The optimal control problem into question is problem $\mathscr{P}_p(0, x_0)$, considered under Assumptions (A1)–(A3), where the parameter function $p \in \Pi$ is uncertain. It is assumed that for some parameter $\hat{p} \in \Pi$ equation (2) with $p = \hat{p}$ reproduces a “real” system, the states of which can be measured (with a measurement errors). As in the previous subsections we denote by $(\hat{x}, \hat{u})$ a reference optimal solution of $\mathscr{P}_{\hat{p}}(0, x_0)$.

Given a natural number N, we denote by $\{t_k\}_{k=0}^N$ the grid with step-size $h = T/N$, that is, $t_k = kh$, $k = 0, \ldots, N$. The MPC algorithm generates in real time an admissible control function, denoted further by u^N. It is applied to the “real” system, that is, (2) with $p = \hat{p}$, resulting in a “real” trajectory x^N.

At time t_k, $k = 0, \ldots, N-1$, the algorithm does the following.

1. Measure the state $x^N(t_k)$ with error e_k, that is, the vector $x_k^0 = x^N(t_k) + e_k$ becomes available.

2. Make a prediction $p_k \in \Pi$ for the time horizon $[t_k, T]$.

3. Find an approximate solution $(\tilde{x}_k, \tilde{u}_k) \in W^{1,1} \times \mathscr{U}_{t_k}$ of the problem $\mathscr{P}_{p_k}(t_k, x_k^0)$.

4. Define the control u^N as $u^N(t) = \tilde{u}_k(t)$ on $(t_k, t_{k+1}]$ and apply to the ”real” system on this interval.

The process continues in the same way as long as $k < N$. The control u^N is called *MPC-generated control* and the corresponding trajectory x^N of the "real" system (2) with $u = u^N$ and $p = \hat{p}$ is called *MPC-generated trajectory*.

Two points are to be clarified. First, the quality of a prediction $p_k \in \Pi$ on $[t_k, T]$ will be measured by the norm $e_k^p := \|p_k - \hat{p}_{[t_k,T]}\|_\infty$. Second, the pair $(\tilde{x}_k, \tilde{u}_k)$ is an approximate solution of problem $\mathscr{P}_{p_k}(t_k, x_k^0)$ in the sense that for some absolutely continuous $\tilde{\lambda}_k$ the triplet $\tilde{y}_k := (\tilde{x}_k, \tilde{\lambda}_k, \tilde{u}_k)$ satisfies the inclusion (approximate optimality conditions)

$$\tilde{z}_k \in \Phi_{(p_k, t_k, x_k^0)}(\tilde{y}_k) \tag{8}$$

with some $\tilde{z}_k \in Z_\tau$. We mention that most of the numerical methods for optimal control give approximations with a small residual $\tilde{z}_k$. The norm $e_k^u := \|\tilde{z}_k\|$ of the residual will be used as a measure of the accuracy of the approximate solution $(\tilde{x}_k, \tilde{u}_k)$ of problem $\mathscr{P}_{p_k}(t_k, x_k^0)$.

3.2 The main theorem

The formulation of the main theorem uses the notations e_k, e_k^p, e_k^u, $\tilde{u}_k$, u^N and x^N introduced in the description of the MPC algorithm. In particular, (x^N, u^N) is the MPC-generated trajectory-control pair, which is compared in the next theorem with the reference optimal open-loop solution $(\hat{x}, \hat{u})$ of the "real" problem $\mathscr{P}_{\hat{p}}(0, x_0)$.

Theorem 3.1. *Let Assumptions (A1)–(A3) be fulfilled. Then there exists numbers N_0, $\delta > 0$, C_1, C_2, and C_3 such that for any natural number $N \geq N_0$, for any sequence of measurement errors $\{e_k\}$, for any sequence of predictions $p_k \in \Pi$ and approximation errors $\{e_k^u\}$ satisfying the conditions*

$$|e_k| + e_k^p + e_k^u \leq \delta, \qquad \|\tilde{u}_k - \hat{u}\|_1 \leq \delta, \quad k = 0, \ldots, N-1,$$

any MPC-generated trajectory-control pair (x^N, u^N) satisfies the estimate

$$\|u^N - \hat{u}\|_1 + \|x^N - \hat{x}\|_{1,1} \leq \begin{cases} C_1 \mathscr{E} & \text{if } \Gamma = \emptyset, \\ C_2\sqrt{\mathscr{E}} + C_3 h & \text{if } \Gamma \neq \emptyset, \end{cases}$$

where

$$\mathscr{E} := \frac{1}{N} \sum_{k=0}^{N-1} (|e_k| + e_k^p + e_k^u)$$

is the averaged error appearing at the MPC steps.

The proof of the theorem is postponed to Section 5. Below in this subsection we discuss the obtained result and the assumptions.

Remark 3.2. *About Assumption (A3).* Assumptions (A1) and (A2) are standard and non-restrictive, although somewhat stronger than necessary, as noted after their formulation. Assumption (A3) has to be explained.

First of all, what is the finite set Γ in the definition of the metric d^* which is involved in (A3) through the definition of strong sub-regularity? This set may depend on the reference optimal control $\hat{u}$ of the unperturbed problem $\mathscr{P}_{\hat{p}}(0, x_0)$. Presumably, it consists of points of discontinuity of $\hat{u}$, but may be larger in order to include points at which the optimal control of a slightly disturbed problem may be discontinuous. Example 3.5 below illustrates this situation. The meaning of the metric d^* is that the distance between two control functions is small in this metric when their values are close to each other, possibly excepting points that are close to the set Γ. This property of the metric with which the Ss-R assumption (A3) is fulfilled is of key importance for that convergence and error analysis of the MPC method.

In the next section we shall provide sufficient conditions under which Assumption (A3) is fulfilled in particular classes of problems with empty or non-empty set Γ.

Remark 3.3. *Discussion on the theorem.* Theorem 3.1 estimates the error of the MPC-generated solution, compared with the optimal solution of the reference (unperturbed) problem, caused by prediction errors $\{e_k^p\}_k$, measurement errors $\{e_k\}_k$, approximation errors $\{e_k^u\}_k$, and the sampling size h. An important point is that the error estimate in the theorem depends on the average error, $\mathscr{E}$, which means that relatively large errors may occasionally appear at some MPC steps without a substantial influence on $\mathscr{E}$.

A similar result as in Theorem 3.1 is obtained in [11] in the case $\Gamma = \emptyset$ (see Subsection 4.1 of the present paper). Here we mention that in [11] a prediction made at the beginning is only used, that is, $p_k = p_0$ for all k. Moreover, the result in [11] only applies to the Euler method for approximately solving the auxiliary problems involved in the MPC algorithm.

Remark 3.4. *About the approximate solution of the auxiliary problems $\mathscr{P}_p(\tau, x_\tau)$.* Finding an approximate solution of the auxiliary problems is a separate issue that we do not address in detail in this paper. Numerical solutions usually involve time-discretization. Discretization methods with first and second order accuracy are known for coercive problems (see Subsection 4.1 for the last term), [8, 10], as well as for affine problems with purely bang-bang optimal controls, [1, 25]. The error in solving the resulting mathematical programming problems comes in addition. The above mentioned results are proved under assumptions that imply strong sub-regularity (for appropriate sets Γ) of the optimality maps associated with the considered problems.

Example 3.5. *Sharpness of the estimate.* Consider the problem

$$\min\{x^1(1) - x^2(1)\},$$

$$\begin{aligned} \dot{x}^1(t) &= p(t)x^2(t), \quad x^1(0) = 0, \\ \dot{x}^2(t) &= u(t), \qquad\quad x^2(0) = 0, \qquad u(t)) \in [-1, 1]. \end{aligned}$$

The reference parameter is $\hat{p} \equiv 1$. The measurements are assumed exact, as well as the solutions of the auxiliary problems at the MPC steps. Thus, in the notations in Theorem 3.1, $\mathscr{E} = h \sum_{k=0}^{N-1} \|p_k - \hat{p}\|_\infty$.

The solution of each problem $\mathscr{P}_p(t_k, \hat{x}(t_k))$ is straightforward: here

$$\tilde{\lambda}_k^2(t) = -1 + \int_t^1 p_k(s)\,\mathrm{d}s, \qquad \tilde{u}_k(t) = -\operatorname{sign} \tilde{\lambda}^2(t).$$

For $\hat{p}$ we have $\hat{\lambda}^2(t) < 0$ for all $t \in (0,1]$, hence $\hat{u}(t) \equiv 1$. For $t = 1$ the control function $\hat{u}$ is not determined by the Pontryagin necessary optimality condition, because $\hat{\lambda}^2(0) = 0$. This does not matter from the control perspective, but suggests to define $\Gamma = \{0\}$ (see Remark 3.2). As it will become obvious in the next section, Assumption (A3) is fulfilled for this problem with this Γ.

Let us fix an arbitrary $\delta > 0$ and consider $h = 1/N$ with $N > 2$ and such that $2h \le \delta$. Define $p_0 = 1 + 2h$ and take all other predictions exact: $p_k = 1, k = 1, \dots, N-1$. Then $\|p_0 - \hat{p}\|_\infty \le 2h \le \delta$. Moreover, $\mathscr{E} = 2h^2$.

On the other hand, we have

$$\tilde{\lambda}_0^2(t_1) = -1 + (1-h)(1+2h) = h(1-2h) > 0.$$

Since $\tilde{\lambda}_0^2$ is linear and $\tilde{\lambda}_0^2(1) = -1$, we obtain that $\tilde{\lambda}_0^2(t) > 0$ on $[0,t_1]$. Hence, $u^N(t) = \tilde{u}_0(t) = -1$ on $[0,t_1]$. Then

$$\|u^N - \hat{u}\|_1 \geq 2h.$$

Consequently,

$$\frac{\|u^N - \hat{u}\|_1}{\sqrt{\mathscr{E}}} \geq \frac{2h}{\sqrt{2h}} \geq \sqrt{2}.$$

Since $\mathscr{E}$ can be arbitrarily small (for small h), the estimation in the theorem is sharp.

4 Sufficient conditions for strong sub-regularity of the optimality map

In this section we present some classes of problems for which Assumption (A3) has a more particular form with a specified set Γ, thus Theorem 3.1 is applicable. In addition, we further discuss the approximation issue mentioned in Remark 3.4.

4.1 The case of coercive problems

Following [9], in this subsection we consider the reference problem $\mathscr{P}_{\hat{p}}(0,x_0)$ under the so-called *coercivity condition*. To formulate it we use the following notational convention: we skip arguments of functions with "hat", shifting the "hat" over the notation of the function, e.g. $\hat{f}_x(t) := f_x(\hat{p}(t),\hat{x}(t),\hat{u}(t))$, $\hat{H}_{xx}(t) := H_{xx}(\hat{p}(t),\hat{x}(t),\hat{\lambda}(t),\hat{u}(t))$, etc. Here f_x and H_{xx} are the derivative of f and the Hessian of H, respectively.

Assumption (B1). There is a constant $c_0 > 0$ such for any $v \in \mathscr{U} - \mathscr{U}$ the inequality

$$\langle g_T''(\hat{x}(T))x(T),x(T)\rangle + \int_0^T \left[\langle \hat{H}_{xx}(t)x(t),x(t)\rangle + 2\langle \hat{H}_{ux}(t)x(t),v(t)\rangle + \langle \hat{H}_{uu}(t)v(t),v(t)\rangle\right] \mathrm{d}t \geq c_0\|v\|_2^2$$

is fulfilled, where x is the (unique) solution of the equation $\dot{x}(t) = \hat{f}_x(t)x(t) + \hat{f}_u(t)v(t)$ with $x(0) = 0$.

It was proved in [9] that Assumption (B1), together with (A1) and (A2), implies (A3) with $\Gamma = \emptyset$, thus in this case the metric in $\mathscr{U}$ is $d^* = \|\cdot\|_\infty$, thus the metric in Y can be taken to be $d^*(y,0) = \|x\|_{1,\infty} + \|\lambda\|_{1,\infty} + \|u\|_\infty$.[1]

Even more, Assumptions (A1), (A2), (B1) imply the stronger property of *Strong metric Regularity* (SR), [14, Sect. 3.7] with respect to the norm $\|x\|_{1,\infty} + \|\lambda\|_{1,\infty} + \|u\|_\infty$ in the space Y. An important fact is, that the property SR is stable with respect to functional perturbations with a sufficiently small Lipschitz constant see, e.g., [14, Proposition 3G.2]). Then it is easy to see that for sufficiently small inaccuracies $|e_k|$, e_k^p and e_k^u all the maps $\Phi_{(p,t_k,x_k^0)}$ that appear in the MPC algorithm are SR, hence Ss-R, with constants that can be chosen uniformly with respect to k. In connection with Remark 3.4, we mention that thanks to the strong regularity of the optimality map (or the uniform strong sub-regularity) one can claim $O(h)$ uniform estimation of e_k^u if the Euler discretization with step size h is used in solving the problems $\mathscr{P}_p(t_k,x_k^0)$ (see [17]), and uniform convergence of the Newton method (see [2]). However, this issue is not at the focus of the present paper and we do not give precise formulations and details.

[1] The terminology of metric regularity was not used in [9] and the control system considered was stationary, but the result was easily extended to the non-stationary case in many subsequent contributions, see e.g. [13].

4.2 The case of affine problems with bang-bang optimal controls

In this subsection, we consider problem $\mathscr{P}_{\hat{p}}(0,x_0)$ to be affine, i.e., the objective integrand, g, and the right-hand side, f, in (2) are both affine with respect to u. The set U is assumed to be a convex compact polyhedron. Using geometric terminology, we denote by V the set of vertices of U, and by E the set of all unit vectors $e \in \mathbb{R}^m$ that are parallel to some edge of U. As usual, we define the so-called switching function $\hat{\sigma} : [0,T] \to \mathbb{R}^m$ by $\hat{\sigma}(t) := \hat{H}_u(t)$. Here and further we use the notational convention from the previous subsection: arguments of functions with "hat" are skipped and the "hat" is shifted over the notation of the functions.

Versions of the following assumption are standard in the literature on affine optimal control problems, see, e.g., [1, 6, 18, 24].

Assumption (C1). There exist numbers $\eta_0 > 0$ and $\mu_0 > 0$ such that if $s \in [0,T]$ is a zero of $\langle \hat{\sigma}, e\rangle$ for some $e \in E$, then

$$|\langle \hat{\sigma}(t), e\rangle| \geq \mu_0 |t-s|,$$

for all $t \in [s-\eta_0, s+\eta_0] \cap [0,T]$.

Assumption (C1) implies that $\hat{u}$ is bang bang and that, in particular, the set

$$\Gamma := \{s \in [0,T] : \langle \hat{\sigma}(s), e\rangle = 0 \text{ for some } e \in E\}$$

is finite. In what follows in this subsection, the metric d^* in Y is defined through this set Γ, see (7). As it will be seen in the proof of Theorem 3.1, the advantage of using this metric instead of the L^1-norm is that $d^*(u,\hat{u})$ being small not only implies that $\|u-\hat{u}\|_1$ is small, but also that $|u(t)-\hat{u}(t)|$ is small except on a small set around the zeros of the switching functions $\langle \hat{\sigma}(t), e\rangle$, $e \in E$. In this sense, u is structurally similar to $\hat{u}$.

Given $\varepsilon \geq 0$, we denote

$$\Sigma(\varepsilon) := [0,T] \setminus (\Gamma + [-\varepsilon, \varepsilon]).$$

We recall the following lemma proved in the recent paper [7].

Lemma 4.1. *[7, Lemma 2] Let Assumption (C2) be fulfilled. Then there exist positive numbers κ and ε such that for every functions $\sigma \in L^\infty$ with $\|\sigma - \hat{\sigma}\|_\infty \leq \varepsilon$ and for every $u \in \mathscr{U}$ satisfying $\sigma(t) + N_U(u(t)) \ni 0$ for a.e. $t \in [0,T]$ it holds that*

$$u(t) = \hat{u}(t) \text{ for a.e. } t \in \Sigma\big(\kappa\|\sigma - \hat{\sigma}\|_\infty\big).$$

With this lemma at hand, we are now ready to establish the following sufficient condition for the fulfillment of Assumption (A3).

Theorem 4.2. *Let problem $\mathscr{P}_{\hat{p}}(0,x_0)$ be affine and let Assumptions (A1), (A2), and (C1) be fulfilled. Then the following statements are equivalent:*

(i) The map $\Phi_{(\hat{p},0,x_0)} : Y \rightrightarrows Z$ is strongly sub-regular at $(\hat{y},0)$ in the single metric d in Y;

(ii) The map $\Phi_{(\hat{p},0,x_0)} : Y \rightrightarrows Z$ is strongly sub-regular at $(\hat{y},0)$ in the metrics d and d^* in Y.

Proof. The implication (ii) $\Rightarrow$ (i) is obvious. Let us prove the converse implication.

Let $\tilde{\alpha}, \tilde{\beta}$ and $\tilde{\kappa}$ be the parameters of Ss-R of $\Phi_{(\hat{p},0,x_0)}$ (in the single metric d in Y). Then

$$\|x-\hat{x}\|_{1,1} + \|\lambda - \hat{\lambda}\|_{1,1} + \|u-\hat{u}\|_1 \leq \tilde{\kappa} d_Z(z,0) \tag{9}$$

for all $y=(x,\lambda,u) \in B(\hat{y};\tilde{\alpha})$ and $z=(\xi,\nu,\eta,\rho) \in B_Z(0,\tilde{\beta})$ satisfying $z \in \Phi_{(\hat{p},0,x_0)}(y)$. Let us fix arbitrarily such a pair (y,z). Define $\sigma : [0,T] \to \mathbb{R}^m$ by $\sigma(t) := \nabla_u H(\hat{p}, y(t)) - \rho$. Clearly $\sigma(t) + N_U(u(t)) \ni 0$ for a.e. $t \in [0,T]$. Moreover, due to the affine structure of the problem, $\nabla_u H(p,y)$ is independent of u, thus we can estimate

$$\begin{aligned} \|\sigma - \hat{\sigma}\|_\infty &\leq \|\nabla_u H(\hat{p},y) - \nabla_u H(\hat{p},\hat{y})\|_\infty + \|\rho\|_\infty \leq L_1(\|x-\hat{x}\|_\infty + \|\lambda - \hat{\lambda}\|_\infty) + \|\rho\|_\infty \\ &\leq L_1 \tilde{\kappa} d_Z(z,0) + d_Z(z,0) =: \tilde{c} d_Z(z,0), \end{aligned}$$

where L_1 is the Lipschitz constant of $\nabla_u H(\hat{p},\cdot)$.

Define

$$\hat{\alpha} = \tilde{\alpha}, \quad \hat{\beta} = \min\left\{\tilde{\beta}, \varepsilon/\tilde{c}\right\}, \quad \hat{\kappa} = \tilde{\kappa} + \tilde{c}\kappa,$$

where ε and κ are the numbers in Lemma 4.1. For the pair (y,z) we additionally assume that $z \in B_Z(0,\hat{\beta})$. Then by Lemma 4.1,

$$u(t) = \hat{u}(t) \text{ for a.e. } t \in \Sigma\big(\kappa \tilde{c} d_Z(z,0)\big),$$

which directly implies $d^*(u,\hat{u}) \leq \kappa \tilde{c} d_Z(z,0)$. Together with (9), we obtain that

$$d^*(y,\hat{y}) \leq \tilde{c}\kappa d_Z(z,0) + \tilde{\kappa} d_Z(z,0) = \hat{\kappa} d_Z(z,0),$$

which proves (ii) with Ss-R constants $\hat{\alpha}$, $\hat{\beta}$, $\hat{\kappa}$. □

A general sufficient condition for strong sub-regularity of the map $\Phi_{(\hat{p},0,x_0)} : Y \rightrightarrows Z$ in the single metric d in Y is given in [24, Theorem 3.1]. It involves the following assumption.

Assumption (C2). There is a constant $c_0 > 0$ such for any $v \in \mathscr{U} - \hat{u}$ the inequality

$$\int_0^T \langle \hat{H}_u(t), v(t)\rangle \,\mathrm{d}t + \langle g_T''(\hat{x}(T))x(T), x(T)\rangle + \int_0^T \left[\langle \hat{H}_{xx}(t)x(t), x(t)\rangle + 2\langle \hat{H}_{ux}(t)x(t), v(t)\rangle\right] \,\mathrm{d}t \geq c_0\|v\|_1^2 \tag{10}$$

is fulfilled, where x is the (unique) solution of the equation $\dot{x}(t) = \hat{f}_x(t)x(t) + \hat{f}_u(t)v(t)$ with $x(0) = 0$.

In [24, Theorem 3.1] it is proved that Assumption (C2), together with (A1), (A2), and the affine structure of the problem, implies metric sub-regularity of the optimality map in the single metric d in Y. In contrast to the L^2 coercivity condition in the previous subsection, Assumption (C2) requires "coercivity" with respect to the L^1-norm. It is well known that Assumption (B1) does not hold for affine problems, see [13, Lemma 3]. Another difference between (B1) and (C2) is that the inequality in (C2) involves not only a quadratic, but also a linear form of v. It is remarkable that alone this linear term can ensure fulfillment of Assumption (C1). Indeed, in [24, Proposition 4.1] it is proved that Assumption (C1) implies the inequality

$$\int_0^T \langle \hat{H}_u(t), v(t)\rangle \,\mathrm{d}t \geq c_1\|v\|_1^2$$

for a constant $c_1 > 0$ and all $v \in \mathscr{U} - \hat{u}$. In particular, if the quadratic form in (10) is nonnegative for $v \in \mathscr{U} - \hat{u}$, then (A1), (A2), (C1) imply (C2), hence also Assumption (A3).

4.3 A numerical example

In this section we illustrate the result obtained in Theorem 3.1 by considering a problem of axisymmetric spacecraft spin stabilization from [28, p. 353]. The transversal angular velocity components ω_1 and ω_2 of the spacecraft satisfy

$$\dot{\omega}_1 = \lambda\omega_2 + \frac{M_d}{J_t},$$
$$\dot{\omega}_2 = -\lambda \sin\omega_1 + \frac{M_c}{J_t},$$

where $\lambda = \frac{J_t - J_3}{J_t} n$, J_t is the spacecraft transversal moment of inertia, J_3 is the spacecraft moment of inertia about the spin axis, n is the spin rate, M_d is the disturbance torque, which can be caused by thruster misalignment, and M_c is the control moment. Rescaling the time $(t \to \lambda t)$, denoting $x_1 = \omega_1$, $x_2 = \omega_2$, $p = \frac{M_d}{J_t}$, adding initial conditions, and considering $u := \frac{M_c}{J_t}$ as a control variable, we reformulate the model as

$$\begin{cases} \dot{x}_1 = x_2 + p, & x_1(0) = 1, \\ \dot{x}_2 = -\sin x_1 + u, & x_2(0) = 1, \\ -a \le u \le a, & \end{cases} \tag{11}$$

where $p(\cdot)$ is a time-dependent parameter and a is a positive constant. The MPC algorithm is applied in [12] to the following optimal control problem with the dynamic (11):

$$\min\left\{ |x(T)|^2 + \alpha \int_0^T (u(t))^2 \, dt \right\},$$

where α is a positive weighting parameter. This problem is coercive in the sense of Assumption (B1), which makes the analysis in [12] possible. Here we consider the alternative objective functional

$$\min\left\{ |x(T)|^2 + \alpha \int_0^T |u(t)| \, dt \right\}, \tag{12}$$

which may be more realistic in case of direct transformation of fuel into force, as in jet engines. The optimal control problem (11)-(12) is not coercive, nor does it fit in the framework of affine problems. However, following [27, Remark 3.3], we transform it to an affine problem by substituting

$$u = u_1 - u_2, \quad |u| = u_1 + u_2, \quad \text{where} \quad u_1, u_2 \in [0, a].$$

Thus, the affine optimal control problem we will consider is

$$\min\left\{ |x(T)|^2 + \alpha \int_0^T [u_1(t) + u_2(t)] \, dt \right\},$$

subject to

$$\begin{cases} \dot{x}_1 = x_2 + p, & x_1(0) = 1, \\ \dot{x}_2 = -\sin x_1 + u_1 - u_2, & x_2(0) = 1, \\ u_1, u_2 \in [0, a]. & \end{cases}$$

We consider the last problem with the specifications $T = 4\pi$, $\alpha = 0.25$, $a = 0.2$, and reference parameter $\hat{p}(t) \equiv 0$. In the MPC simulation, the measurement error e_k is sampled randomly from a uniform distribution, with $|e_k| \le 0.1$. The parameter $p(\cdot)$ is piecewise constant on the uniform mesh of 3200 points in $[0, T]$; its values in every subinterval are chosen randomly in the interval $[-0.05, 0.05]$ with uniform distribution. For solving the auxiliary problems $\mathscr{P}_p(t_k, x_k^0)$ we use the Euler discretization scheme, which provides an error e_k^u of order $O(h)$ see, e.g., [1, 24]); recall that $e_k^u := \|\tilde{z}_k\|$ and that $\tilde{z}_k$ is the residual in (8).

We run the MPC algorithm with different mesh sizes N. Using the notations in Theorem 3.1, we consider the quantity

$$RE = \frac{||u^N - \hat{u}||_1 + ||x^N - \hat{x}||_{1,1}}{\sqrt{\frac{1}{N}\sum_{k=0}^{N-1}(|e_k| + e_k^p + h) + h}}, \tag{13}$$

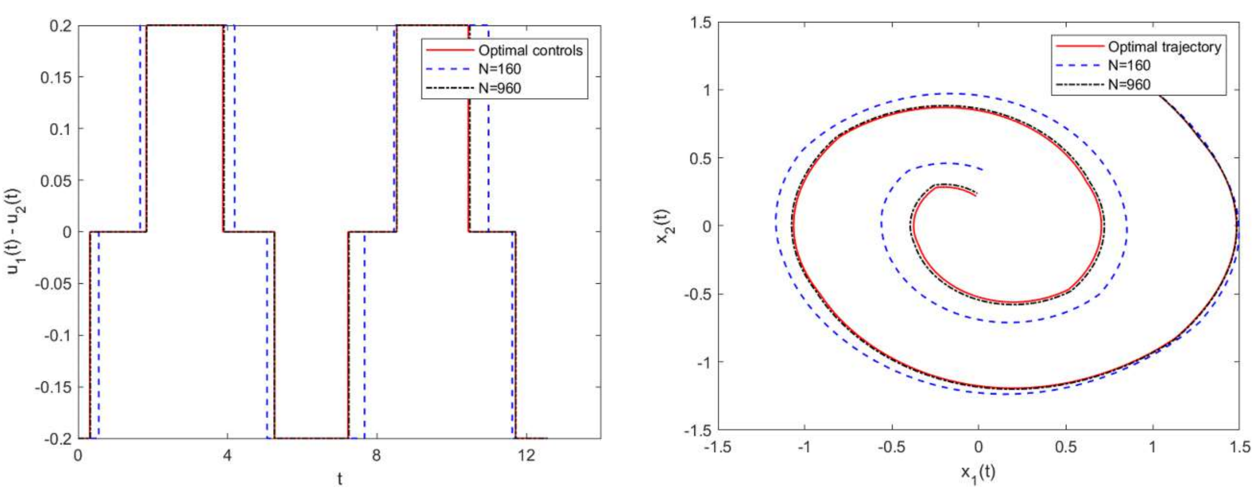
which represents the relative error in the MPC-generated solution (x^N, u^N). According to the error estimate in Theorem 3.1, the quantity RE should be bounded. The numerical experiment confirms this, as can be seen in Table 1. Moreover, the result suggest that the value RE stays away from zero when N increases, which indicates that the estimate in Theorem 3.1 is sharp for this example.

We also observe in Table 1 that the objective values for the MPC-generated solutions decrease when N increases, which is to be expected because of the more frequent measurements.

In Figure 1, we compare the obtained MPC-generated controls and the corresponding trajectories with the open-loop solution $\hat{u} = \hat{u}_1 - \hat{u}_2$. The auxiliary controls $\hat{u}_1$ and $\hat{u}_2$ are of bang-bang type, while the resulting optimal control u in problem (11)-(12) also takes value zero. The MPC-generated control u^N differs from the optimal open-loop one in small intervals around the switching points of the latter, which is consistent with the choice of the metric in d^* in case of affine problems.

N	160	320	480	640	800	960
Obj. val.	1.3221	0.7907	0.7204	0.6490	0.6274	0.6183
RE	0.6039	0.3205	0.2177	0.1245	0.1445	0.1346

Table 1: Objective values and relative errors *RE* of the MPC-generated solutions with different mesh sizes.



(a) Controls

(b) Trajectories

Figure 1: Optimal open-loop control and trajectory, and the MPC generated solutions with N=160 and N=960.

5 Proof of the main theorem

In this section we use the notations introduced in Sections 2 and 3. In addition, $y \in Y_\tau$ we define

$$d^*_\tau(y,\hat{y}) := d^*(\tilde{y},\hat{y}), \qquad \text{where } \tilde{y}(t) = \begin{cases} \hat{y}(t) & \text{for } t \in [0,\tau), \\ y(t) & \text{for } t \in [\tau,T]. \end{cases}$$

Proposition 5.1. *Let assumptions (A1)–(A3) be fulfilled. Then there exist numbers $\delta_0 > 0$, $\alpha_0 > 0$, $\beta_0 > 0$, and c_0 such that for every $\tau \in [0,T)$, $x_\tau \in \mathbb{R}^n$ and $p \in \Pi$ satisfying*

$$|x_\tau - \hat{x}(\tau)| \le \delta_0, \qquad \|p - \hat{p}\|_\infty \le \delta_0, \tag{14}$$

and for every $y = (x,\lambda,u) \in Y_\tau$ with $\|u - \hat{u}\|_1 \le \alpha_0$ and $z_\tau \in \Phi_{(p,\tau,x_\tau)}(y) \cap \boldsymbol{B}_{Z_\tau}(0;\beta_0)$ it holds that

$$d^*_\tau(y,\hat{y}) \le c_0\left(\|z_\tau\| + \|p - \hat{p}\|_\infty + |x_\tau - \hat{x}(\tau)|\right).$$

Proof. We shall fix the numbers $\delta_0 > 0$, $\alpha_0 > 0$, $\beta_0 > 0$ and c_0 below, in a way that they depend only on $\hat{\alpha}$, $\hat{\beta}$, $\hat{\kappa}$, L and T and may be viewed as constants. The numbers c_1, c_2, ..., that will appear later will also be appropriate constants in the same sense.

Let us fix arbitrarily $\tau \in [0,T)$, $x_\tau \in \mathbb{R}^n$, and $p \in \Pi$ satisfying (14), along with $y_\tau = (x(\cdot),\lambda(\cdot),u(\cdot)) \in B(\hat{y};\alpha_0)$ and $z_\tau = (\xi,\nu,\eta,\pi,\rho) \in \boldsymbol{B}_{Z_\tau}(0;\beta_0)$ such that $z_\tau \in \Phi_{(p,\tau,x_\tau)}(y_\tau)$. We define

$$\tilde{p}(t) = \begin{cases} \hat{p}(t) & \text{for } t \in [0,\tau), \\ p(t) & \text{for } t \in [\tau,T], \end{cases} \qquad \tilde{u}(t) = \begin{cases} \hat{u}(t) & \text{for } t \in [0,\tau), \\ u(t) & \text{for } t \in [\tau,T]. \end{cases}$$

Obviously $\|\tilde{p} - \hat{p}\|_\infty = \|p - \hat{p}\|_\infty$, $\|\tilde{u} - \hat{u}\|_1 = \|u - \hat{u}\|_1$, and $d^*_0(\tilde{u},\hat{u}) = d^*_\tau(u,\hat{u})$. Similarly, we extend ξ, η and ρ as zero on $[0,\tau)$, denoting the resulting elements of Z_0 by $\tilde{\xi}$, $\tilde{\eta}$, $\tilde{\rho}$. Moreover, we define $\tilde{x}$ as the (unique) solution on $[0,T]$ of the equations

$$\dot{\tilde{x}} = f(\tilde{p},\tilde{x},\tilde{u}) - \tilde{\xi}, \quad \tilde{x}(\tau) = x_\tau + \nu,$$

and the function $\tilde{\lambda}$ as the solution of

$$-\dot{\tilde{\lambda}} = \nabla_x H(\tilde{p},\tilde{x},\tilde{\lambda},\tilde{u}) - \eta, \quad \tilde{\lambda}(T) = \nabla \tilde{g}_T(\tilde{x}(T)) + \pi.$$

Notice that $\tilde{x}(t) = x(t)$ and $\tilde{\lambda}(t) = \lambda(t)$ for $t \in [\tau,T]$.

Let us estimate $\|\tilde{y} - \hat{y}\|$. Using Assumption (A1) and the Grönwall inequality we obtain that

$$\|\tilde{x} - \hat{x}\|_\infty \le c_1\left(\|\tilde{p} - \hat{p}\|_\infty + \|\tilde{\xi}\|_1 + |x_\tau - \hat{x}(\tau)| + |\nu|\right) \le c_2\left(\delta_0 + \|\xi\|_1 + \delta_0 + |\nu|\right) \le c_3(\delta_0 + \|z_\tau\|).$$

From here one can also estimate $\|\dot{\tilde{x}} - \dot{\hat{x}}\|_1$, which gives

$$\|\tilde{x} - \hat{x}\|_{1,1} \le c_4(\delta_0 + \|z_\tau\|).$$

Similarly we estimate

$$\|\tilde{\lambda} - \hat{\lambda}\|_{1,1} \le c_5(\delta_0 + \|z_\tau\|).$$

Moreover,

$$\|\tilde{u} - \hat{u}\|_1 = \|u - \hat{u}\|_1 \le \alpha_0.$$

The last three estimates imply that

$$\|\tilde{y}-\hat{y}\| \leq c_6(\delta_0+\alpha_0+\beta_0) \leq \hat{\alpha},$$

provided that the positive numbers δ_0, α_0 and β_0 are chosen sufficiently small.

Now we estimate the residual $r := (r_\xi, r_\nu, r_\eta, r_\pi, r_\rho) \in Z_0$ which $\tilde{y} := (\tilde{x}, \tilde{\lambda}, \tilde{u})$ gives in $\Phi_{(\hat{p},0,x_0)}$. We have

$$\begin{aligned}
\|r_\xi\|_1 &= \left\|f(\hat{p},\tilde{x},\tilde{u}) - f(\tilde{p},\tilde{x},\tilde{u}) + \tilde{\xi}\right\|_1 \leq TL\|p-\hat{p}\|_\infty + \|\xi\|_1, \\
|r_\nu| &= |\tilde{x}(0)-x_0| \leq c_1(\|p-\hat{p}\|_\infty + |x_\tau - \hat{x}(\tau)| + |\nu|), \\
\|r_\eta\|_1 &= \left\|\nabla_x H(\hat{p},\tilde{y}) - \nabla_x H(\tilde{p},\tilde{y}) + \tilde{\eta}\right\|_1 \leq TL\|p-\hat{p}\|_\infty + \|\eta\|_1, \\
|r_\pi| &= \pi|, \\
\|r_\rho\|_\infty &= \|\nabla_u H(\hat{p},\tilde{y}) - \nabla_u H(\tilde{p},\tilde{y}) + \tilde{\rho}\|_\infty \leq L\|p-\hat{p}\|_\infty + \|\rho\|_\infty.
\end{aligned}$$

Summarizing, we obtain that

$$\|r\| \leq c_7\left(\|z_\tau\| + \|p-\hat{p}\|_\infty + |x_\tau - \hat{x}(\tau)|\right).$$

We can choose δ_0 and β_0 smaller if needed so that $\|r\| \leq \hat{\beta}$. Due to Assumption (A3) we have that

$$d_0^*(\tilde{y},\hat{y}) \leq \hat{\kappa} c_7\left(\|z_\tau\| + \|p-\hat{p}\|_\infty + |x_\tau - \hat{x}(\tau)|\right).$$

Since $d_\tau^*(\tilde{y},\hat{y}) \leq d_0^*(\tilde{y},\hat{y})$ and $\tilde{y} = y$ on $[\tau, T]$, we obtain the desired result with $c_0 = \hat{\kappa} c_7$. □

An alternative way to prove the last proposition is first to show that Assumption (A3) holds for all maps $\Phi_{(\hat{p},\tau,\hat{x}(\tau))}$ with $\tau \in [0,T)$, and then to apply Proposition 2.2.

Now we continue with the proof of Theorem 3.1.

Let δ_0, α_0, β_0 and c_0 be the constants from Proposition 5.1. Define the following constants: M is number of elements of Γ (equals zero if $\Gamma = \emptyset$) and

$$D := \operatorname{diam}(U), \qquad \bar{C}_1 := c_0 L T e^{LT}, \qquad \bar{C}_2 := 2DLe^{TL}\sqrt{c_0 T M}, \qquad \bar{C}_3 := 6MDLe^{LT}. \tag{15}$$

Let the numbers N_0 and $\delta > 0$ be defined in such a way that

$$\hat{\delta} + \delta \leq \delta_0, \qquad \delta \leq \alpha_0, \qquad \delta \leq \beta_0,$$

where $\hat{\delta} := \bar{C}_1\delta + \bar{C}_2\sqrt{\delta} + \bar{C}_3 h$, which is obviously possible. Moreover, denote

$$\mathscr{E}_i := |e_i| + e_i^p + e_i^u, \;\; i = 0,\ldots,N-1.$$

Since for any $i \in \{0,\ldots,N-1\}$ the triplet $y = \tilde{y}_i$ satisfies (8), we shall apply Proposition 5.1 with $y = \tilde{y}_i$, $\tau = t_i$ $x_\tau = x_i^0 = x^N(t_i) + e_i$, $p = p_i$, $z_\tau = \tilde{z}_i$. We have

$$\|p_i - \hat{p}\|_\infty \leq e_i^p \leq \delta \leq \delta_0,$$

$$\|\tilde{u}_i - \hat{u}\|_1 \leq \delta \leq \alpha_0,$$

$$\|\tilde{z}_i\| = e_i^u \leq \delta \leq \beta_0.$$

Proposition 5.1 gives

$$d^*_{t_i}(\tilde{u}_i, \hat{u}) \le c_0(|e_i| + e^p_i + e^u_i) = c_0\mathscr{E}_i, \tag{16}$$

provided that

$$|x^N(t_i) + e_i - \hat{x}(t_i)| \le \delta_0. \tag{17}$$

Let us fix an arbitrary $k \in \{1, \ldots, N-1\}$ and denote

$$d_i := \|\tilde{u}_i - \hat{u}\|_{L^\infty(t_i,t_{i+1})}, \qquad \Delta(t) := |x^N(t) - \hat{x}(t)|, \qquad \Delta_i = \Delta(t_i),\ i = 0, \ldots, k.$$

Assume inductively that

$$\Delta_k \le \hat{\delta} \quad \text{and} \quad d^*_{t_i}(\tilde{u}_i, \hat{u}) \le c_0\mathscr{E}_i,\ i = 0, \ldots, k. \tag{18}$$

For $k = 0$ we have $\Delta_0 = |x^N(0) - \hat{x}(0)| = 0$. Thus (17) is fulfilled because $|e_0| \le \delta \le \delta_0$. The second inequality in (18) is fulfilled due to (16), thus the inductive assumption is fulfilled for $k = 0$.

Due to Assumption (A1) and the construction of u^N in the MPC method, we have for $t \in [t_k, t_{k+1}]$

$$\Delta(t) \le \Delta_k + \int_{t_k}^t |f(\hat{p}(s), x^N(s), \tilde{u}_k(s)) - f(\hat{p}(s), \hat{x}(s), \hat{u}(s))| \, ds \le \Delta_k + \int_{t_k}^t (L\Delta(s) + L|\tilde{u}_k(s) - \hat{u}(s)|) \, ds.$$

Using the Grönwall inequality we obtain that

$$\Delta(t) \le e^{Lh}(\Delta_k + hLd_k).$$

Applied to $\Delta_{k+1} = \Delta(t_{k+1})$, this recursive inequality implies in a standard way that

$$\Delta_{k+1} \le hL\Big(e^{(k+1)hL}d_0 + e^{khL}d_1 + \ldots + e^{hL}d_k\Big) \le e^{TL}Lh\sum_{i=0}^k d_i. \tag{19}$$

The key part of the proof is to estimate $\sum_{i=0}^k d_i$. Let us denote

$$\begin{aligned} \bar{K} &:= \{i \in \{0, \ldots, k\} : |\tilde{u}_i(t) - \hat{u}(t)| \le d^*_{t_i}(\tilde{u}_i, \hat{u}) \text{ for a.e. } t \in [t_i, t_{i+1}]\}, \\ K &:= \{0, \ldots, k\} \setminus \bar{K}. \end{aligned}$$

Then

$$d_i \le \begin{cases} D & \text{for } i \in K, \\ d^*_{t_i}(\tilde{u}_i, \hat{u}) & \text{for } i \in \bar{K}. \end{cases} \tag{20}$$

Denoting by s the number of elements of K, we have due to the inductive assumption (18), that

$$\sum_{i=0}^k d_i = \sum_{i \in K} d_i + \sum_{i \in \bar{K}} d_i \le sD + \sum_{i \in \bar{K}} d^*_{t_i}(\tilde{u}_i, \hat{u}) \le sD + \sum_{i \in \bar{K}} c_0\mathscr{E}_k \le sD + \frac{c_0 T}{h}\mathscr{E}. \tag{21}$$

Let us assume that $\Gamma \ne \emptyset$, that is, $M > 0$. The definition of the metric d^* in (7) implies that for each $i \in K$ there exists $t \in (t_i, t_{i+1})$ such that

$$\text{dist}(t, \Gamma) \le d^*_{t_i}(\tilde{u}_i, \hat{u}). \tag{22}$$

Let $m(i)$ be the minimal natural number (also including 0) such that

$$((t_i, t_{i+1}) + h[-m(i), m(i)]) \cap \Gamma \ne \emptyset.$$

Then in the case $m(i) > 0$ we have that

$$((t_i, t_{i+1}) + h[-m(i)+1, m(i)-1]) \cap \Gamma = \emptyset,$$

hence

$$(t + h[-m(i)+1, m(i)-1]) \cap \Gamma = \emptyset.$$

Due to (22) we obtain that

$$d^*_{t_i}(\tilde{u}_i, \hat{u}) \geq h(m(i)-1), \quad i \in K. \tag{23}$$

Denote by l_j, $j = 0, 1, \ldots, N$ the number of those indexes $i \in K$ for which $m(i) = j$. The following relations are apparently satisfied:

$$\begin{aligned} &0 \leq l_0 \leq M, \\ &0 \leq l_j \leq 2M, \quad j = 1, \ldots, N, \\ &\sum_{j=0}^{N} l_j = s. \end{aligned} \tag{24}$$

Then, having in mind (23),

$$\sum_{i \in K} d^*_{t_i}(\tilde{u}_i, \hat{u}) \geq h \sum_{i \in K} \max\{0, m(i)-1)\} \geq h\Big(0.l_0 + 0.l_1 + \sum_{j=2}^{N} (j-1) l_j\Big). \tag{25}$$

The minimum of the sum in the right-hand side with respect to $\{l_j\}$ subject to the relations around (24) is attained at

$$l_0 = M, \; l_j = 2M, \; j = 1, \ldots, r, \quad l_{r+1} = s - (M + 2Mr),$$

where $r = \left[\frac{s-M}{2M}\right]$ and $[a]$ means the integer part of a. Substituting l_j in (25) we obtain that

$$\sum_{i \in K} d^*_{t_i}(\tilde{u}_i, \hat{u}) \geq 2hM \sum_{j=2}^{r} (j-1) = hMr(r-1) \geq hM\left(\left[\frac{s}{2M}\right] - 2\right)^2.$$

From here and the second inequality in (18) we obtain that

$$\left([\frac{s}{2M}] - 2\right)^2 \leq \frac{c_0}{Mh} \sum_{i=0}^{k} \mathscr{E}_i = \frac{c_0 T}{Mh^2} \mathscr{E},$$

hence

$$s \leq 6M + \frac{2}{h}\sqrt{c_0 T M \mathscr{E}}.$$

From (21) we obtain that

$$h \sum_{i=0}^{k} d_i \leq D\Big(6Mh + 2\sqrt{c_0 T M}\sqrt{\mathscr{E}}\Big) + c_0 T \mathscr{E}, \tag{26}$$

which combined with (19) and (15) gives

$$\Delta_{k+1} \leq \bar{C}_1 \mathscr{E} + \bar{C}_2 \sqrt{\mathscr{E}} + \bar{C}_3 h. \tag{27}$$

This inequality was obtained in the case $M > 0$. However, in the case $M = 0$ the first term in the final inequality in (21) is missing and the analysis simplifies, resulting in the same estimation for Δ_{k+1} but with $M = 0$, which implies $\bar{C}_2 = \bar{C}_3 = 0$.

Now, in order to verify the inductive assumption (18) we observe that $\mathscr{E} \le \delta$, hence from (27)

$$\Delta_{k+1} \le \bar{C}_1\delta + \bar{C}_2\sqrt{\delta} + \bar{C}_3 h = \hat{\delta},$$

thus the first inequality in (18) is satisfied for $k+1$. The second inequality in (18) for $k+1$ follows from (16), which is fulfilled for $i = k+1$ because (17) holds:

$$|x^N(t_{k+1}) + e_{k+1} - \hat{x}(t_{k+1})| \le \Delta_{k+1} + |e_{k+1}| \le \hat{\delta} + \delta \le \delta_0.$$

This completes the induction step. As a result we have obtained that for any $t \in [0,T]$ if $t \in [t_k, t_{k+1}]$ then

$$\Delta(t) \le \Delta_{k+1} \le \bar{C}_1\mathscr{E} + \bar{C}_2\sqrt{\mathscr{E}} + \bar{C}_3 h. \tag{28}$$

Now we estimate

$$\begin{aligned} \|u^N - \hat{u}\|_1 &= \sum_{i=0}^{N-1} \int_{t_i}^{t_{i+1}} |u^N(t) - \hat{u}(t)|\,\mathrm{d}t \le \sum_{i=0}^{N-1} h\|u^N - \hat{u}\|_{L^\infty(t_i,t_{i+1})} = \sum_{i=0}^{N-1} h\|\tilde{u}_i - \hat{u}\|_{L^\infty(t_i,t_{i+1})} = h\sum_{i=0}^{N-1} d_i \\ &\le D\Big(6Mh + 2\sqrt{c_0TM}\sqrt{\mathscr{E}}\Big) + c_0T\mathscr{E}, \end{aligned} \tag{29}$$

where in the last inequality we use (26) for $k = N-1$. Finally, we have

$$\|\dot{x}^N - \dot{\hat{x}}\|_1 \le \int_0^T |f(\hat{p}(t), x^N(t), u^N(t)) - f(\hat{p}(t), \hat{x}(t), \hat{u}(t))|\,\mathrm{d}t \le LT\|\Delta\|_C + L\|u^N - \hat{u}\|_1.$$

Combining this inequality with (28) and (29), and considering separately the case $M = 0$, we obtain existence of constants C_1, C_2 and C_3 for which the claim of the theorem holds.

References

[1] W. Alt, U. Felgenhauer, M. Seydenschwanz. Euler discretization for a class of nonlinear optimal control problems with control appearing linearly. *Compututaional Optimization and Applications*, **69**:825–856, 2018.

[2] F.J. Aragón, A.L. Dontchev, M. Gaydu, M.H. Geoffroy, V.M. Veliov. *SIAM Journal on Control and Optimization*, **49**(2):339–362], 2011.

[3] *Encycopedia of Sistems abd Control*. J. Baillieul and T. Samad (Eds.). Springer Nature, 2020.

[4] J.F. Bonnans, A. Shapiro. *Perturbation analysis of optimization problems*. Springer, 2000.

[5] R. Cibulka, A.L. Dontchev, A.Y. Kruger. Strong metric subregularity of mappings in variational analysis and optimization. *Journal of Mathematical Analysis and Application*, **457**:1247–1282, 2018.

[6] A. Domínguez Corella., M. Quincampoix., V.M. Veliov.: Strong bi-metric regularity in affine optimal control problems. To appear in *Pure and Applied Functional Analysis*. Available at https://orcos.tuwien.ac.at/research/

[7] A. Domínguez Corella., V.M. Veliov. Hölder Regularity in Bang-Bang Type Affine Optimal Control Problems. To appear in *Large-scale scientific computing*. Available at https://orcos.tuwien.ac.at/research/

[8] A. L. Dontchev. An a priori estimate for discrete approximations in nonlinear optimal control. *SIAM J. Control Optim.*, **34**:1315–1328, 1996.

[9] A.L. Dontchev, W.W. Hager. Lipschitzian stability in nonlinear control and optimization. *SIAM J. Control Optim.*, **31**:569–603, 1993.

[10] A.L. Dontchev, W.W. Hager, V.M. Veliov. Second-order Runge-Kutta approximations in control constrained optimal control. *SIAM J. Numerical Anal.*, **38**(1):202–226, 2000.

[11] A.L. Dontchev, I.V. Kolmanovsky, M. I. Krastanov, V.M. Veliov. Approximating open-loop and closed-loop optimal control by model predictive control. *European Control Conference (ECC)*, 2020, DOI: 10.23919/ECC51009.2020.9143615.

[12] A.L. Dontchev, I.V. Kolmanovsky, M.I. Krastanov, V.M. Veliov, P.T. Vuong. Approximating optimal finite horizon feedback by model predictive control. *Systems & Control Letters*, **139**:104666, May 2020.

[13] A.L. Dontchev, M.I. Krastanov, V.M. Veliov. On the existence of Lipschitz continuous optimal feedback control. *Vietnam Journal of Mathematic*, **47**(3):579–597, 2019.

[14] A.L. Dontchev, R.T. Rockafellar. Characterizations of Lipschitz stability in nonlinear programming *Mathematical programming with data perturbations*, 65–82, Lecture Notes in Pure and Appl. Math., 195, Dekker, New York, 1998.

[15] A.L. Dontchev, R.T. Rockafellar. Regularity and conditioning of solution mappings in variational analysis. *Set-Valued Analysis*, **12**:79–109, 2004.

[16] A.L. Dontchev, R.T. Rockafellar. *Implicit Functions and Solution Mappings: A View from Variational Analysis. Second edition*. Springer, New York , 2014.

[17] A.L Dontchev, V.M. Veliov. Metric regularity under approximations. *Control and Cybernetics*, **38**(4):1283–1303 , 2009.

[18] U. Felgenhauer. On stability of bang-bang type controls. SIAM J. of Control and Optimization, **41**(6), 1843–1867, 2003.

[19] G. Grimm, M. Messina, S. Tuna, and A.R. Teel. Nominally robust model predictive control with state constraints. *IEEE Trans. Automatic Control*, **52**(10):1856–1870, 2007.

[20] L. Grüne and V. Palma. On the benefit of re-optimization in optimal control under perturbations. Proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems, pp. 439–446, 2014.

[21] L. Grüne, J. Panek. *Nonlinear model predictive control*. Springer, 2011.

[22] D. Klatte, B. Kummer. *Nonsmooth equations in optimization*. Kluwer Academic Publisher, 2002.

[23] W. Langson, I. Chryssochoos, S.V. Raković, and D.Q. Mayne. Robust model predictive control usin tubes. *Automatica*, **40**:125–133, 2004.

[24] N.P. Osmolovskii, V.M. Veliov. Metric sub-regularity in optimal control of affine problems with free end state. ESAIM Control Optim. Calc. Var., **26**, Paper No. 47 (2020)

[25] A. Pietrus, T. Scarinci, V.M. Veliov. High order discrete approximations to Mayer's problems for linear systems. *SIAM J. Control Optim.*, **56**(1):102–119, 2018.

[26] S. V. Raković, W. S. Levine, Eds. *Handbook of Model Predictive Control.* Birkhäuser 2018.

[27] S.P Sethi, G. L. Thompson. Optimal control theory. Applications to management science and economics *Kluwer Academic Publishers, Boston, MA*, 2000.

[28] B. Wie. *Space vehicle dynamics and control*, AIAA , 1998.

[29] I.J. Wolf, W. Marquardt. Fast NMPC schemes for regulatory and economic NMPC–A review. *Journal of Process Control*, **44**:162–183, 2016.

Chapter 4

Stability in Affine Optimal Control Problems Constrained by Elliptic Partial Differential Equation

This chapter consists of the paper:

- Domínguez Corella, Alberto and Jork, Nicolai and Veliov, Vladimir M. Stability in affine optimal control problems constrained by semilinear elliptic partial differential equations.

It is conditionally accepted (after a minor revision) in ESAIM: Control, Optimisation and Calculus of Variations. The thesis author gave most of the ideas and proofs in the paper, and wrote the majority of the paper; he was only helped with technical details and proof-reading.

Stability in affine optimal control problems constrained by semilinear elliptic partial differential equations*

Alberto Domínguez Corella† Nicolai Jork ‡ Vladimir Veliov§

August 14, 2022

Abstract

This paper investigates stability properties of affine optimal control problems constrained by semilinear elliptic partial differential equations. This is done by studying the so called metric subregularity of the set-valued mapping associated with the system of first order necessary optimality conditions. Preliminary results concerning the differentiability of the functions involved are established, especially the so-called switching function. Using this ansatz, more general nonlinear perturbations are encompassed, and under weaker assumptions, than the ones previously considered in the literature on control constrained elliptic problems. Finally, the applicability of the results is illustrated with some error estimates for the Tikhonov regularization.

1 Introduction

We consider the following optimal control problem

$$\min_{u\in\mathcal{U}}\left\{\int_\Omega \Big[w(x,y)+s(x,y)u\Big]\,dx\right\}, \tag{1}$$

subject to

$$\left\{\begin{array}{rcll} -\operatorname{div}\big(A(x)\nabla y\big)+d(x,y) &=& \beta(x)u & \text{in} \quad \Omega \\ A(x)\nabla y\cdot\nu+b(x)y &=& 0 & \text{on} \quad \partial\Omega. \end{array}\right. \tag{2}$$

The set $\Omega\subset\mathbb{R}^n$ is a bounded domain with Lipschitz boundary, where $n\in\{2,3\}$. The unit outward normal vector field on the boundary $\partial\Omega$, which is single valued a.e. in $\partial\Omega$, is denoted by ν. The control set is given by

$$\mathcal{U}:=\{u:\Omega\to\mathbb{R}\ \text{ measurable}:\ \ b_1(x)\le u(x)\le b_2(x)\ \text{ for a.e. } x\in\Omega\},$$

where b_1 and b_2 are bounded measurable functions satisfying $b_1(x)\le b_2(x)$ for a.e. $x\in\Omega$. The functions $w:\Omega\times\mathbb{R}\to\mathbb{R}$, $s:\Omega\times\mathbb{R}\to\mathbb{R}$, $d:\Omega\times\mathbb{R}\to\mathbb{R}$, $\beta:\Omega\to\mathbb{R}$ and $b:\partial\Omega\to\mathbb{R}$ are real-valued and measurable, and $A:\Omega\to\mathbb{R}^{n\times n}$ is a measurable matrix-valued function.

There are many motivations for studying stability of solutions, in particular for error analysis of numerical methods, see e.g., [30, 31]. Most of the stability results for elliptic control problems are obtained under a second order growth condition (analogous to the classical Legendre-Clebsch condition). For literature concerning this type of problems, the reader is referred to [18, 21, 22, 24, 25, 35] and the references therein. In optimal control problems like (1)–(2), where the control appears linearly (hence, called affine problems)

*This research was supported by the Austrian Science Foundation (FWF) under grant I 4571-N.
†Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, alberto.corella@tuwien.ac.at
‡Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, nicolai.jork@tuwien.ac.at
§Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, vladimir.veliov@tuwien.ac.at

this growth condition does not hold. The so-called bang-bang solutions are ubiquitous in this case, see [4, 9, 10]. To give an account of the state of art in stability of bang-bang problems, we mention the works [1, 28, 29, 33, 37] on optimal control of ordinary differential equations. Results for optimization problems constrained by partial differential equations have been gaining relevance in recent years, see [5, 8, 9, 10, 12, 34]. However, its stability has been only investigated in a handful of papers, see e.g., [12, 32, 34]. From these works, we mention here particularly [34], where the authors consider linear perturbations in the state and adjoint equations for a similar problem with Dirichlet boundary condition. They use the so-called structural assumption (a growth assumption satisfied near the jumps of the control) on the adjoint variable. This assumption has been widely used in the literature on bang-bang control of ordinary differential equations in a somewhat different form.

The investigations of stability properties of optimization problems, in general, are usually based on the study of similar properties of the corresponding system of necessary optimality conditions. The first order necessary optimality conditions for problem (1)–(2) can be recast as a system of two elliptic equations (primal and adjoint) and one variational inequality (representing the minimization condition of the associated Hamiltonian), forming together a *generalized equation*, that is, an inclusion involving a set-valued mapping called *optimality mapping*. The concept of *strong metric subregularity*, see [11, 16], of set-valued mappings has shown to be efficient in many applications especially ones related to error analysis, see [2]. This also applies to optimal control problems of ordinary differentials equations, see e.g., [15, 28].

In the present paper we investigate the strong metric subregularity property of the optimality mapping associated with problem (1)–(2). We present sufficient conditions for strong subregularity of this mapping on weaker assumptions than the ones used in literature, see Section 6 for precise details. The structural assumption in [34] is weakened and more general perturbations are considered. Namely, perturbations in the variational inequality, appearing as a part of the first order necessary optimality conditions, are considered; which are important in the numerical analysis of ODE and PDE constrained optimization problems. Moreover, nonlinear perturbations are investigated, which provides a framework for applications, as illustrated with an estimate related to the Tikhonov regularization. The concept of linearization is employed in a functional frame in order to deal with nonlinearities. The needed differentiability of the control-to-adjoint mapping and the switching function (see Section 3) is proved, and the derivatives are used to obtain adequate estimates needed in the stability results. Finally, we consider nonlinear perturbations in a general framework. We propose the use of the compact-open topology to have a notion of "closeness to zero" of the perturbations. In our particular case this topology can me metrized, providing a more "quantitative" notion. Estimates in this metric are obtained in Section 5.

2 Preliminaries

The euclidean space $\mathbb{R}^s$ is considered with its usual norm, denoted by $|\cdot|$. As usual, for $p \in [1, \infty)$, we denote by $L^p(\Omega)$ the space of all measurable p-integrable functions $\psi : \Omega \to \mathbb{R}^s$ with the norm

$$|\psi|_{L^p(\Omega)} := \Big(\sum_{i=1}^{s} \int_\Omega |\psi_i(x)|^p \, dx \Big)^{\frac{1}{p}}.$$

The space $L^\infty(\Omega)$ consists of all measurable essentially bounded functions $\psi : \Omega \to \mathbb{R}^s$ with the norm

$$|\psi|_{L^\infty(\Omega)} := \operatorname*{ess\,sup}_{x\in\Omega} |\psi(x)|.$$

We denote by $C(\bar{\Omega})$ the space of continuous functions on Ω that can be extended continuously to $\bar{\Omega}$ equipped with the L^∞-norm. We denote by $H^1(\Omega)$ the space of functions $\psi \in L^2(\Omega)$ with weak derivatives in $L^2(\Omega)$ endowed with its usual norm. The space $H^1(\Omega) \cap C(\bar{\Omega})$ is endowed with the norm

$$|\psi|_{H^1(\Omega)\cap C(\bar{\Omega})} := |\psi|_{H^1(\Omega)} + |\psi|_{C(\bar{\Omega})}.$$

A function $\psi : \Omega \times \mathbb{R} \to \mathbb{R}$ is said to be Carathéodory if $\psi(\cdot, y)$ is measurable for every $y \in \mathbb{R}$, and $\psi(x, \cdot)$ is continuous for a.e. $x \in \Omega$. A function $\psi : \Omega \times \mathbb{R} \to \mathbb{R}$ is said to be locally Lipschitz uniformly in the first variable if for each $M > 0$ there exists $L > 0$ such that

$$|\psi(x, y_2) - \psi(x, y_1)| \le L|y_2 - y_1|$$

for a.e. $x \in \Omega$ and all $y_1, y_2 \in [-M, M]$. In order to abbreviate notation, we define $f, g : \Omega \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by

$$f(x, y, u) := \beta(x)u - d(x, y) \quad \text{and} \quad g(x, y, u) := w(x, y) + s(x, y)u.$$

The following assumption is supposed to hold throughout the remainder of the paper. It ensures that the mathematical objects related to problem (1)–(2) that we consider are well defined. Assumption 1 is quite standard in the literature, see the book [39].

Assumption 1. *The following statements are assumed to hold.*

(i) *The set $\Omega \subset \mathbb{R}^n$ is a bounded Lipschitz domain. The matrix $A(x)$ is symmetric for a.e. x in Ω, and there exists $\alpha > 0$ such that $\xi \cdot A(x)\xi \geq \alpha|\xi|^2$ for a.e. x in Ω and all $\xi \in \mathbb{R}^n$.*

(ii) *The functions w, s and d are Carathéodory, twice differentiable with respect to the second variable, and their second derivatives are locally Lipschitz uniformly in the first variable.*

(iii) *The functions $A, \beta, b, d(\cdot, 0), d_y(\cdot, 0), w_y(\cdot, 0)$ and $s_y(\cdot, 0)$ are measurable and bounded.*

(iv) *The function $d_y(\cdot, y)$ is nonnegative a.e. in Ω for all $y \in \mathbb{R}$. The function b is nonnegative a.e. in $\partial\Omega$ and $|b|_{L^\infty(\partial\Omega)} > 0$.*

Items (i) and (iv) of Assumption 1 ensure that the partial differential equations appearing in this paper have unique solutions in the space $H^1(\Omega) \cap L^\infty(\Omega)$.

2.1 The elliptic operator

We consider the set $D(\mathcal{L})$ of all functions $y \in H^1(\Omega) \cap L^\infty(\Omega)$ for which there exists $h \in L^2(\Omega)$ such that

$$\int_\Omega A(x)\nabla y \cdot \nabla\varphi \, dx + \int_{\partial\Omega} b(x)y\varphi \, ds(x) = \int_\Omega h\varphi \, dx \quad \forall \varphi \in H^1(\Omega). \tag{3}$$

As usual, ds denotes the Lebesgue surface measure. It is easy to see that for each $y \in D(\mathcal{L})$ there exists a unique element $h \in L^2(\Omega)$ such that (3) holds. We define the operator $\mathcal{L} : D(\mathcal{L}) \to L^2(\Omega)$ by assigning each $y \in D(\mathcal{L})$ to the function $h \in L^2(\Omega)$ satisfying (3). By definition, a function $y \in H^1(\Omega) \cap L^\infty(\Omega)$ belongs to $D(\mathcal{L})$ if, and only if, it is the weak solution of the linear elliptic partial differential equation

$$\begin{cases} -\operatorname{div}\big(A(x)\nabla y\big) &= h \quad \text{in} \quad \Omega, \\ A(x)\nabla y \cdot \nu + b(x)y &= 0 \quad \text{on} \quad \partial\Omega \end{cases}$$

for some $h \in L^2(\Omega)$. The following lemma is of trivial nature.

Lemma 2.1. *The set $D(\mathcal{L})$ is a linear subspace of $H^1(\Omega) \cap L^\infty(\Omega)$. Moreover, the operator $\mathcal{L} : D(\mathcal{L}) \to L^2(\Omega)$ is a well defined linear mapping.*

If $D(\mathcal{L})$ is endowed with the norm of $L^2(\Omega)$, then $\mathcal{L}$ is an unbounded operator from $D(\mathcal{L})$ to $L^2(\Omega)$. Since $A(x)$ is symmetric for a.e. $x \in \Omega$, by (3) we have

$$\int_\Omega \mathcal{L}y\varphi \, dx = \int_\Omega y\mathcal{L}\varphi \, dx \tag{4}$$

for all $y, \varphi \in D(\mathcal{L})$, the so-called integration by parts formula.

Remark 2.2. If $\partial\Omega$ is of class $C^{1,1}$, A is Lipschitz in $\bar{\Omega}$, and b is Lipschitz and positive in $\partial\Omega$, then

$$D(\mathcal{L}) = \{y \in H^2(\Omega) : A(\cdot)\nabla y \cdot \nu + b(\cdot)y = 0\},$$

and $\mathcal{L}y = -\operatorname{div}\big(A(\cdot)\nabla y\big)$ for all $y \in D(\mathcal{L})$, see [19, Theorem 2.4.2.6].

The following lemma shows the inclusion $D(\mathcal{L}) \subset C(\bar{\Omega})$. Its proof can be found in [39, Theorem 4.7] and follows the arguments in [4, 38].

Lemma 2.3. *Let $\alpha \in L^\infty(\Omega)$ be nonnegative and $h \in L^2(\Omega)$. There exists a unique function $y \in D(\mathcal{L})$ such that*

$$\mathcal{L}y + \alpha(\cdot)y = h \tag{5}$$

and this function belongs to $C(\bar{\Omega})$. Moreover, for each $r > n/2$ there exists a positive number c such that

$$|y|_{H^1(\Omega)\cap C(\bar{\Omega})} \le c|h|_{L^r(\Omega)}$$

for all $\alpha \in L^\infty(\Omega)$ nonnegative, $y \in D(\mathcal{L})$, and $h \in L^2(\Omega) \cap L^r(\Omega)$ satisfying (5).

The following technical lemma can be deduced from Lemma 2.3, see the proof of [10, Lemma 3.4]. Its use in optimal control of elliptic partial differential equations dates from the paper [9, Lemma 2.6]. It has shown to be useful for diverse estimates, see [9, 34].

Lemma 2.4. *There exists a positive number c such that*

$$|y|_{L^2(\Omega)} \le c|h|_{L^1(\Omega)}$$

for all $\alpha \in L^\infty(\Omega)$ nonnegative, $y \in D(\mathcal{L})$ and $h \in L^2(\Omega)$ satisfying (5).

The proof of the next result can be found in [7, Theorem 2.11] in the case of a Dirichlet problem, see also [20, Lemma 6.8]. Here we adapt the argument below Theorem 2.1 in [6, p. 618].

Lemma 2.5. *Let $\alpha \in L^\infty(\Omega)$ be nonnegative, $\{h_m\}_{m=1}^\infty$ be a sequence in $L^2(\Omega)$ and $h \in L^2(\Omega)$. For each $m \in \mathbb{N}$, let $y_m \in C(\bar{\Omega})$ be the unique function satisfying $\mathcal{L}y_m + \alpha(\cdot)y_m = h_m$, and let $y \in C(\bar{\Omega})$ be the unique function satisfying of $\mathcal{L}y + \alpha(\cdot)y = h$. If $h_m \rightharpoonup h$ weakly in $L^2(\Omega)$, then $y_m \to y$ in $C(\bar{\Omega})$.*

Proof. Let $p \in (2n/(n+2), n/(n-1))$. Then $W^{1,p}(\Omega)$ is compactly embedded in $L^2(\Omega)$ and consequently, by Schauder's Theorem, $L^2(\Omega)$ is compactly embedded in $W^{1,p}(\Omega)^*$. By the latter compact embedding, every weakly convergent sequence in $L^2(\Omega)$ converges also in $W^{1,p}(\Omega)^*$ to the same limit. Define $\mathcal{K} : L^2(\Omega) \to C(\bar{\Omega})$ by $\mathcal{K}h := y$, where $y \in C(\bar{\Omega})$ is the unique function satisfying $\mathcal{L}y + \alpha(\cdot) = h$. The result follows from [27, Theorem 3.14], since that theorem asserts that the linear operator $\mathcal{K}$ is continuous from $L^2(\Omega)$ endowed with the norm of $W^{1,p}(\Omega)^*$ to $C(\bar{\Omega})$. □

Remark 2.6. Using the definitions of the set $D(\mathcal{L})$ and the operator $\mathcal{L}$, we can write in a shorter way the partial differential equations involved in this paper. For example, given $u \in \mathcal{U}$, to say that y belongs to $D(\mathcal{L})$ and satisfies $\mathcal{L}y + d(\cdot, y) = \beta(\cdot)u$ is equivalent to say that y belongs to $H^1(\Omega) \cap L^\infty(\Omega)$ and satisfies the weak formulation of (2), that is

$$\int_\Omega A(x)\nabla y \cdot \nabla\varphi \, dx + \int_\Omega d(x,y)\varphi \, dx + \int_{\partial\Omega} b(x)y\varphi \, ds(x) = \int_\Omega \beta(x)u\varphi \, dx$$

for all $\varphi \in H^1(\Omega)$. This weak formulation makes sense since, by *(ii)* and *(iii)* of Assumption 1, for any $y \in L^\infty(\Omega)$, the function $d(\cdot, y)$ belongs to $L^\infty(\Omega)$.

2.2 The control model

Having in mind Remark 2.6, given a function $u \in \mathcal{U}$ we say that $y_u \in D(\mathcal{L})$ is the associated state to $u \in \mathcal{U}$ if

$$\mathcal{L}y_u = f(\cdot, y_u, u). \tag{6}$$

The following proposition shows that the mapping $u \to y_u$ from $\mathcal{U}$ to $D(\mathcal{L})$ is well defined. Its proof can be found in the standard literature; it follows from [39, Theorem 4.8], see also [39, p. 212].

Proposition 2.7. *For each $u \in \mathcal{U}$ there exists a unique state $y_u \in D(\mathcal{L})$ associated with $u \in \mathcal{U}$. Moreover, $\{y_u : u \in \mathcal{U}\}$ is a bounded subset of $H^1(\Omega) \cap C(\bar{\Omega})$ and for each $r > n/2$ there exists $c > 0$ such that*

$$|y_{u_2} - y_{u_1}|_{H^1(\Omega)\cap C(\bar{\Omega})} \le c|u_2 - u_1|_{L^r(\Omega)}$$

for all $u_1, u_2 \in \mathcal{U}$.

We call the function $\mathcal{G} : \mathcal{U} \to H^1(\Omega) \cap C(\bar{\Omega})$ given by $\mathcal{G}(u) := y_u$ the control-to-state mapping. The functional $\mathcal{J} : \mathcal{U} \to \mathbb{R}$ given by

$$\mathcal{J}(u) := \int_\Omega g(x, y_u, u)\, dx$$

is called the objective functional of problem (1)–(2).

Definition 2.8. Let $\bar{u}$ belong to $\mathcal{U}$.

(i) We say that $\bar{u}$ is a global solution of problem (1)–(2) if $\mathcal{J}(\bar{u}) \le \mathcal{J}(u)$ for all $u \in \mathcal{U}$.

(ii) We say that $\bar{u}$ is a local solution of problem (1)–(2) if there exists $\varepsilon_0 > 0$ such that $\mathcal{J}(\bar{u}) \le \mathcal{J}(u)$ for all $u \in \mathcal{U}$ with $|u - \bar{u}|_{L^1(\Omega)} \le \varepsilon_0$.

(iii) We say that $\bar{u}$ is a strict local solution of problem (1)–(2) if there exists $\varepsilon_0 > 0$ such that $\mathcal{J}(\bar{u}) < \mathcal{J}(u)$ for all $u \in \mathcal{U}$ with $u \ne \bar{u}$ and $|u - \bar{u}|_{L^1(\Omega)} \le \varepsilon_0$.

Under Assumption 1, problem (1)–(2) has at least one global solution. The proof is routine and can be obtained by standard arguments; namely, taking a minimizing sequence and using the weak compactness of $\mathcal{U}$ in $L^2(\Omega)$.

Lemma 2.9. *Problem (1)–(2) has at least one global solution.*

In order to make notation simpler, from now on we fix a local solution $\bar{u} \in \mathcal{U}$ of problem (1)–(2). We call the function $H : \Omega \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, given by

$$H(x, y, p, u) := g(x, y, u) + p f(x, y, u),$$

the Hamiltonian of problem (1)–(2). Given $u \in \mathcal{U}$, we say that $p_u \in D(\mathcal{L})$ is the costate associated with $u \in \mathcal{U}$ if

$$\mathcal{L} p_u = H_y(\cdot, y_u, p_u, u).$$

The following proposition shows that the mapping $u \to p_u$ from $\mathcal{U}$ to $D(\mathcal{L})$ is well defined. We give the proof of this elementary result because it seems not to be explicitly stated in the literature.

Proposition 2.10. *For each $u \in \mathcal{U}$ there exists a unique costate $p_u \in D(\mathcal{L})$ associated with $u \in \mathcal{U}$. Moreover, $\{p_u : u \in \mathcal{U}\}$ is a bounded subset of $H^1(\Omega) \cap C(\bar{\Omega})$ and for each $r > n/2$ there exist $c > 0$ such that*

$$|p_{u_2} - p_{u_1}|_{H^1(\Omega) \cap C(\bar{\Omega})} \le c |u_2 - u_1|_{L^r(\Omega)}$$

for all $u_1, u_2 \in \mathcal{U}$.

Proof. The existence and uniqueness follows from Lemma 2.3. Given $u \in \mathcal{U}$, the function p_u satisfies

$$\mathcal{L} p_u + d_y(\cdot, y_u) p_u = g_y(\cdot, y_u, u).$$

By $(ii), (iii)$ and (iv) of Assumption 1, for each $u \in \mathcal{U}$, the function $d_y(\cdot, y_u)$ is nonnegative and belongs to $L^\infty(\Omega)$. By (ii) and (iii) of Assumption 1, for each $u \in \mathcal{U}$ the function $g_y(\cdot, y_u, u)$ belongs to $L^\infty(\Omega)$. Furthermore, since by Proposition 2.7 the set $\{y_u : u \in \mathcal{U}\}$ is bounded in $C(\bar{\Omega})$, there exists $M_1 > 0$ such that

$$|g_y(\cdot, y_u, u)|_{L^\infty(\Omega)} \le M_1$$

for all $u \in \mathcal{U}$. By Lemma 2.3, there exists a positive number c_1 such that for all $u \in \mathcal{U}$

$$|p_u|_{H^1(\Omega) \cap C(\bar{\Omega})} \le c_1 |g_y(\cdot, y_u, u)|_{L^\infty(\Omega)}.$$

Thus, $M_2 := c_1 M_1$ is a bound for the set $\{p_u : u \in \mathcal{U}\}$ in $H^1(\Omega) \cap C(\bar{\Omega})$. Let $u_1, u_2 \in \mathcal{U}$ and $r > n/2$. We have then

$$\mathcal{L}(p_{u_2} - p_{u_1}) + d_y(\cdot, y_{u_2})(p_{u_2} - p_{u_1}) = H_y(\cdot, y_{u_2}, p_{u_1}, u_2) - H_y(\cdot, y_{u_1}, p_{u_1}, u_1).$$

By Lemma 2.3, there exists a positive number c_2 (independent of u_1 and u_2) such that

$$|p_{u_2} - p_{u_1}|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq c_2 |H_y(\cdot, y_{u_2}, p_{u_1}, u_2) - H_y(\cdot, y_{u_1}, p_{u_1}, u_1)|_{L^r(\Omega)}.$$

By (ii) of Assumption 1 and the boundedness of the set $\{p_u : u \in \mathcal{U}\}$ in $C(\bar{\Omega})$, there exists $L > 0$ such that

$$|H_y(\cdot, y_{u_2}, p_{u_1}, u_2) - H_y(\cdot, y_{u_1}, p_{u_1}, u_1)| \leq L\Big(|y_{u_2} - y_{u_1}| + |u_2 - u_1|\Big) \quad \text{a.e. in } \Omega.$$

Consequently,

$$\begin{aligned} |p_{u_2} - p_{u_1}|_{H^1(\Omega)\cap C(\bar{\Omega})} &\leq c_2 L\big(|y_{u_1} - y_{u_2}|_{L^r(\Omega)} + |u_1 - u_2|_{L^r(\Omega)}\big) \\ &\leq c_2 L\Big((\operatorname{meas}\Omega)^{\frac{1}{r}} |y_{u_2} - y_{u_1}|_{L^\infty(\Omega)} + |u_2 - u_1|_{L^r(\Omega)}\Big). \end{aligned}$$

By Proposition 2.7, there exists a constant $c_3 > 0$ (independent of u_1 and u_2) such that

$$|y_{u_2} - y_{u_1}|_{C(\bar{\Omega})} \leq c_3 |u_2 - u_1|_{L^r(\Omega)}.$$

Thus,

$$|p_{u_2} - p_{u_1}|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq c_2 L\big(1 + c_3 (\operatorname{meas}\Omega)^{\frac{1}{r}}\big)|u_2 - u_1|_{L^r(\Omega)}.$$

The estimate follows defining $c := c_2 L\big(1 + c_3(\operatorname{meas}\Omega)^{\frac{1}{r}}\big)$. □

We call the function $\mathcal{S} : \mathcal{U} \to H^1(\Omega) \cap C(\bar{\Omega})$ given by $\mathcal{S}(u) := p_u$ the control-to-adjoint mapping. The following proposition gives us another useful estimate; it can be easily proved employing Lemma 2.4 and the argument in the proof of [39, Theorem 4.16].

Proposition 2.11. *There exists $c > 0$ such that*

$$|y_{u_2} - y_{u_1}|_{L^2(\Omega)} + |p_{u_2} - p_{u_1}|_{L^2(\Omega)} \leq c|u_2 - u_1|_{L^1(\Omega)}$$

for all $u_1, u_2 \in \mathcal{U}$.

We close this subsection with the following result.

Proposition 2.12. *Let $\{u_m\}_{m=1}^\infty$ be a sequence in $\mathcal{U}$ and $u \in \mathcal{U}$. If $u_m \rightharpoonup u$ weakly in $L^2(\Omega)$, then $y_{u_m} \to y_u$ and $p_{u_m} \to p_u$ in $C(\bar{\Omega})$.*

Proof. We prove only the convergence $p_{u_m} \to p_u$ in $C(\bar{\Omega})$, the convergence $y_{u_m} \to y_u$ in $C(\bar{\Omega})$ is analogous. Let $\{p_{u_{m_k}}\}_{k=1}^\infty$ be an arbitrary subsequence of $\{p_{u_m}\}_{m=1}^\infty$. By the compact embedding $H^1(\Omega) \hookrightarrow L^2(\Omega)$, there exists a subsequence of $\{p_{u_{m_k}}\}_{k=1}^\infty$, denoted in the same way, and $p \in L^2(\Omega)$ such that $p_{u_{m_k}} \to p$ in $L^2(\Omega)$. Since $y_{u_{m_k}} \to y_u$ in $C(\bar{\Omega})$, one can deduce that

$$H_y(\cdot, y_{u_{m_k}}, p_{u_{m_k}}, u_{m_k}) \rightharpoonup H_y(\cdot, y_u, p, u) \quad \text{weakly in } L^2(\Omega).$$

By Lemma 2.5, we have $p_{u_{m_k}} \to p_u$ in $C(\bar{\Omega})$. The result follows, since every subsequence of $\{p_{u_m}\}_{m=1}^\infty$ has a further subsequence that converges to p_u in $C(\bar{\Omega})$. □

3 Differentiability of the mappings involved

In this section, we prove some preliminary results concerning the differentiability of the control-to-state mapping, the control-to-adjoint mapping and the switching mapping (to be defined later). Some of these properties are well known for the control-to-state mapping; see, e.g., [5, 9, 10, 34, 39]. Nevertheless, we require more specific estimates than the ones in the literature. The differentiability of the control-to-adjoint mapping and the switching mapping has not been studied before in the literature on elliptic control-constrained problems, therefore we devote this section to obtain appropriate estimates needed in the study of stability in the next section. In the sequel, we treat differentiability by means of Gâteaux differentials, as they provide a very natural setting that adjusts in a very versatile way to our purposes.

3.1 The state and adjoint mappings

We begin this subsection recalling the definition of Gâteaux differential, see [17, pp.2-4] or [23, p.171]. Let Y be a Banach space and $\mathcal{F} : \mathcal{U} \to Y$ a mapping. Given $u \in \mathcal{U}$ and $v \in \mathcal{U} - u$, if the limit

$$d\mathcal{F}(u; v) := \lim_{\varepsilon \to 0^+} \frac{\mathcal{F}(u + \varepsilon v) - \mathcal{F}(u)}{\varepsilon}$$

exists in Y, we say that $\mathcal{F}(u; v)$ is the Gâteaux differential of $\mathcal{F}$ at u in the direction v. Note that by convexity of $\mathcal{U}$, $u + \varepsilon v$ belongs to $\mathcal{U}$ for every $u \in \mathcal{U}$, $v \in \mathcal{U} - u$ and $\varepsilon \in [0, 1]$.

Recall that $\bar{u} \in \mathcal{U}$ is a fixed solution of problem (1)–(2). As it is well-known, the Gâteaux differential of the control-to-state mapping at $\bar{u}$ is related to the linearization of the system equation around $\bar{u}$. Bearing this in mind, given $v \in L^2(\Omega)$, we denote by z_v the unique[1] solution of the equation

$$\mathcal{L} z_v = f_y(\cdot, y_{\bar{u}}, \bar{u}) z_v + f_u(\cdot, y_{\bar{u}}, \bar{u}) v. \tag{7}$$

The proof of the following estimate can be found in the standard literature, see the proof of [39, Theorem 4.17] for the case of a Neumann boundary problem (the proof is the same for Robin or Dirichlet boundary). It can also be deduced by the same arguments given in the proof of Proposition 3.2.

Proposition 3.1. *For each $r > n/2$ there exists $c > 0$ such that*

$$|y_u - y_{\bar{u}} - z_{u-\bar{u}}|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq c|u - \bar{u}|^2_{L^r(\Omega)} \qquad \forall u \in \mathcal{U}.$$

One of the first things that can be deduced from Proposition 3.1 is the differentiability of the control-to-state mapping $\mathcal{G}$. Given $v \in L^2(\Omega)$ satisfying $\bar{u} + v \in \mathcal{U}$, the Gâteaux differential of the control-to-state mapping $\mathcal{G}$ at $\bar{u}$ in the direction v exists and is given by $d\mathcal{G}(\bar{u}; v) = z_v$. Moreover, one can prove that $\mathcal{G}$ is of class C^2. This is a standard application of the Implicit Function Theorem to the function $\mathcal{F} : D(\mathcal{L}) \times L^r(\Omega) \to L^r(\Omega)$ given by $\mathcal{F}(y, u) := \mathcal{L}y + d(\cdot, y) - \beta(\cdot)u$, where $r > n/2$; see [7, Theorem 2.12] for details in the Dirichlet boundary case.

In order to study the Gâteaux differential of the control-to-adjoint mapping we introduce the following notations. Given $v \in L^2(\Omega)$, we denote by q_v the unique[2] solution of the equation

$$\mathcal{L} q_v = H_{yy}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) z_v + H_{yp}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) q_v + H_{yu}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) v. \tag{8}$$

The following estimate is concerned with the differentiability of the control-to-adjoint mapping. To the best of our knowledge, this result does not appear in the literature; therefore we present its proof, although it is standard.

Proposition 3.2. *For each $r > n/2$ there exists $c > 0$ such that*

$$|p_u - p_{\bar{u}} - q_{u-\bar{u}}|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq c|u - \bar{u}|^2_{L^r(\Omega)} \qquad \forall u \in \mathcal{U}.$$

Proof. Given $u \in \mathcal{U}$, we define $\psi_u : \Omega \to \mathbb{R}^4$ by $\psi_u(x) := (x, y_u(x), p_u(x), u(x))$. For each $u \in \mathcal{U}$, we denote by $\tilde{q}_{u-\bar{u}}$ the unique solution of the equation

$$\mathcal{L}\tilde{q}_{u-\bar{u}} = H_{yy}(\psi_{\bar{u}})(y_u - y_{\bar{u}}) + H_{yp}(\psi_{\bar{u}})\tilde{q}_{u-\bar{u}} + H_{yu}(\psi_{\bar{u}})(u - \bar{u}).$$

Let $u \in \mathcal{U}$ and $r > n/2$ be arbitrary. Using the Taylor Theorem and (ii)-(iii) of Assumption 1, one can find $\alpha_1, \alpha_2, \alpha_3 \in L^\infty(\Omega)$ such that

$$\begin{aligned} H_y(\psi_u) =& H_y(\psi_{\bar{u}}) + H_{yy}(\psi_{\bar{u}})(y_u - y_{\bar{u}}) + H_{yp}(\psi_{\bar{u}})(p_u - p_{\bar{u}}) + H_{yu}(\psi_{\bar{u}})v \\ &+ \alpha_1(\cdot)(y_u - y_{\bar{u}})^2 + \alpha_2(\cdot)(y_u - y_{\bar{u}})(p_u - p_{\bar{u}}) + \alpha_3(\cdot)(y_u - y_{\bar{u}})v, \end{aligned}$$

[1] The uniqueness follows from Lemma 2.3, and the fact that equation (7) can be rewritten as

$$\mathcal{L} z_v + d_y(\cdot, y_{\bar{u}}) z_v = \beta(\cdot) v.$$

[2] The uniqueness follows from Lemma 2.3, and the fact that equation (8) can be rewritten as

$$\mathcal{L} q_v + d_y(\cdot, y_{\bar{u}}) q_v = H_{yy}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) z_v + H_{yu}(\cdot, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) v.$$

where $v = u - \bar{u}$. Hence

$$\mathcal{L}(p_u - p_{\bar{u}} - \tilde{q}_v) = H_{yp}(\psi_{\bar{u}})(p_u - p_{\bar{u}} - \tilde{q}_v) + \Big[\alpha_1(\cdot)(y_u - y_{\bar{u}}) + \alpha_2(\cdot)(p_u - p_{\bar{u}}) + \alpha_3(\cdot)v\Big](y_u - y_{\bar{u}}).$$

By Lemma 2.3, Proposition 2.7 and Proposition 2.10, there exists $c_1 > 0$ such that

$$|p_u - p_{\bar{u}} - \tilde{q}_v|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq c_1 |v|^2_{L^r(\Omega)}.$$

Now,

$$\mathcal{L}(\tilde{q}_v - q_v) = H_{yy}(\psi_{\bar{u}})(y_u - y_{\bar{u}} - z_v) + H_{yp}(\psi_{\bar{u}})(\tilde{q}_v - q_v).$$

By Lemma 2.3 and Proposition 3.1, there exists $c_2 > 0$ such that

$$|\tilde{q}_v - q_v|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq c_2 |v|^2_{L^r(\Omega)}.$$

Finally, by the triangle inequality

$$|p_u - p_{\bar{u}} - q_v|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq |p_u - p_{\bar{u}} - \tilde{q}_v|_{H^1(\Omega)\cap C(\bar{\Omega})} + |\tilde{q}_v - q_v|_{H^1(\Omega)\cap C(\bar{\Omega})}.$$

The result follows taking $c := c_1 + c_2$. □

Given $v \in L^\infty(\Omega)$ satisfying $\bar{u} + v \in \mathcal{U}$, the Gâteaux differential of the control-to-adjoint mapping $\mathcal{S}$ at $\bar{u}$ in the direction v exists and is given by $d\mathcal{S}(\bar{u}; v) = q_v$. It is worth mentioning that the map $\mathcal{S}$ is of class C^2, this can be seen applying the Implicit Function Theorem to the function $\mathcal{H} : D(\mathcal{L}) \times L^r(\Omega) \to L^r(\Omega)$ given by $\mathcal{H}(p, u) := \mathcal{L}p - H_y(\cdot, y_u, p, u)$, where $r > n/2$.

We now state further properties concerning the mappings $v \to z_v$ and $v \to q_v$.

Proposition 3.3. *The following statements hold.*

(i) For each $r > n/2$ there exists a positive number c such that

$$|z_v|_{H^1(\Omega)\cap C(\bar{\Omega})} + |q_v|_{H^1(\Omega)\cap C(\bar{\Omega})} \leq c|v|_{L^r(\Omega)} \quad \forall v \in L^2(\Omega) \cap L^r(\Omega).$$

(ii) There exists a positive number c such that

$$|z_v|_{L^2\Omega} + |q_v|_{L^2\Omega} \leq c|v|_{L^1(\Omega)} \quad \forall v \in L^2(\Omega).$$

(iii) Let $\{v_k\}_{k=1}^\infty$ be a sequence in $L^2(\Omega)$ and $v \in L^2(\Omega)$. If $v_k \rightharpoonup v$ weakly in $L^2(\Omega)$, then $z_{v_k} \to z_v$ and $q_{v_k} \to q_v$ in $C(\bar{\Omega})$.

Proof. Items (i) and (ii) follow from Lemma 2.3 and 2.4, respectively. Item (iii) follows from Lemma 2.5. □

3.2 The switching mapping

Let us begin this subsection recalling the first order necessary condition (Pontryagin principle in integral form) for problem (1)–(2). If $u \in \mathcal{U}$ is a local solution of problem (1)–(2), then

$$\int_\Omega \Big[s(x, y_u) + \beta(x)p_u\Big](w - u)\, dx \geq 0 \quad \forall w \in \mathcal{U}. \tag{9}$$

The variational inequality (9) motivates the following definition. For each $u \in \mathcal{U}$, define

$$\sigma_u := s(\cdot, y_u) + \beta(\cdot)p_u.$$

The mapping $\mathcal{Q} : \mathcal{U} \to L^\infty(\Omega)$ given by $\mathcal{Q}(u) := \sigma_u$ is called the switching mapping. Given $v \in L^2(\Omega)$, we define the linearization

$$\pi_v := H_{uy}(\cdot, y_{\bar{u}}, p_{\bar{u}})z_v + H_{up}(\cdot, y_{\bar{u}}, p_{\bar{u}})q_v.$$

This definition is justified by the following estimate.

Proposition 3.4. *For each $r > n/2$ there exists $c > 0$ such that*

$$|\sigma_u - \sigma_{\bar{u}} - \pi_{u-\bar{u}}|_{L^\infty(\Omega)} \leq c|u - \bar{u}|^2_{L^r(\Omega)} \qquad \forall u \in \mathcal{U}.$$

Proof. Given $u \in \mathcal{U}$, we define $\psi_u : \Omega \to \mathbb{R}^3$ by $\psi_u(x) := (x, y_u(x), p_u(x))$. For each $u \in \mathcal{U}$, we denote

$$\tilde{\pi}_{u-\bar{u}} := H_{uy}(\psi_{\bar{u}})(y_u - y_{\bar{u}}) + H_{up}(\psi_{\bar{u}})(p_u - p_{\bar{u}}).$$

Let $u \in \mathcal{U}$ and $r > n/2$ be arbitrary, and abbreviate $v = u - \bar{u}$. Using the Taylor Theorem and *(ii)*-*(iii)* of Assumption 1, one can find $\alpha_1, \alpha_2 \in L^\infty(\Omega)$ such that

$$\begin{aligned} H_u(\psi_u) =& H_u(\psi_{\bar{u}}) + H_{uy}(\psi_{\bar{u}})(y_u - y_{\bar{u}}) + H_{up}(\psi_{\bar{u}})(p_u - p_{\bar{u}}) \\ &+ \alpha_1(\cdot)(y_u - y_{\bar{u}})^2 + \alpha_2(\cdot)(y_u - y_{\bar{u}})(p_u - p_{\bar{u}}). \end{aligned}$$

Therefore, by Proposition 2.7 and 2.10, there exists $c_1 > 0$ such that

$$|\sigma_u - \sigma_{\bar{u}} - \tilde{\pi}_v|_{L^\infty(\Omega)} \leq c_1 |v|^2_{L^r(\Omega)}.$$

Now,

$$|\tilde{\pi}_v - \pi_v|_{L^\infty(\Omega)} \leq |H_{uy}(\cdot, y_{\bar{u}}, p_{\bar{u}})(y_u - y_{\bar{u}} - z_v) + H_{up}(\cdot, y_{\bar{u}}, p_{\bar{u}})(q_u - q_{\bar{u}} - q_v)|_{L^\infty(\Omega)}.$$

Hence, by Proposition 3.1 and 3.2, there exists $c_2 > 0$ such that

$$|\tilde{\pi}_v - \pi_v|_{L^\infty(\Omega)} \leq c_2 |v|^2_{L^r(\Omega)}.$$

Finally, by the triangle inequality,

$$|\sigma_u - \sigma_{\bar{u}} - \pi_v|_{L^\infty(\Omega)} \leq |\sigma_u - \sigma_{\bar{u}} - \tilde{\pi}_v|_{L^\infty(\Omega)} + |\tilde{\pi}_v - \pi_v|_{L^\infty(\Omega)}.$$

The result follows defining $c := c_1 + c_2$. □

Proposition 3.4 yields immediately that the Gâteaux differential of the switching mapping $\mathcal{Q}$ at $\bar{u}$ in any direction $v \in \mathcal{U} - \bar{u}$ exists and is given by $d\mathcal{Q}(\bar{u}; v) = z_v$.

One of the important features of the mapping $v \to \pi_v$ is the following.

Proposition 3.5. *For all $v \in L^2(\Omega)$, we have*

$$\int_\Omega \pi_v v \, dx = \int_\Omega \Big[H_{yy}(x, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) z_v^2 + 2H_{uy}(x, y_{\bar{u}}, p_{\bar{u}}, \bar{u}) z_v v \Big] \, dx.$$

Proof. In order to simplify notation, we write $\psi_{\bar{u}}(x) := (x, y_{\bar{u}}(x), p_{\bar{u}}(x), \bar{u}(x))$ for each $x \in \Omega$. Let $v \in L^2(\Omega)$ be arbitrary. By the integration by parts formula (4), we get

$$\begin{aligned} \int_\Omega H_{up}(\psi_{\bar{u}}) q_v v \, dx &= \int_\Omega \big(\mathcal{L} z_v + d_y(x, y_{\bar{u}}) z_v\big) q_v \, dx = \int_\Omega \big(\mathcal{L} q_v + d_y(x, y_{\bar{u}}) q_v\big) z_v \, dx \\ &= \int_\Omega \big(H_{yy}(\psi_{\bar{u}}) z_v + H_{uy}(\psi_{\bar{u}}) v\big) z_v = \int_\Omega \Big[H_{yy}(\psi_{\bar{u}}) z_v^2 + H_{uy}(\psi_{\bar{u}}) z_v v \Big] \, dx. \end{aligned}$$

The result follows since

$$\int_\Omega \pi_v v \, dx = \int_\Omega H_{uy}(\psi_{\bar{u}}) z_v v \, dx + \int_\Omega H_{up}(\psi_{\bar{u}}) q_v v \, dx.$$

□

We give further properties of the mapping $v \to \pi_v$ in the next proposition, its proof follows trivially from Proposition 3.3.

Proposition 3.6. *The following statements hold.*

(i) *For each $r > n/2$ there exists a positive number c such that*

$$|\pi_v|_{L^\infty(\Omega)} \le c|v|_{L^r(\Omega)} \quad \forall v \in L^2(\Omega) \cap L^r(\Omega).$$

(ii) *There exists a positive number c such that*

$$|\pi_v|_{L^2(\Omega)} \le c|v|_{L^1(\Omega)} \quad \forall v \in L^2(\Omega).$$

(iii) *Let $\{v_k\}_{k=1}^\infty$ be a sequence in $L^2(\Omega)$ and $v \in L^2(\Omega)$. If $v_k \rightharpoonup v$ weakly in $L^2(\Omega)$, then $\pi_{v_k} \to \pi_v$ in $L^\infty(\Omega)$.*

Proposition 3.5 motivates the following definition. For each $v \in L^2(\Omega)$, define

$$\Lambda(v) := \int_\Omega \Big[H_{yy}(x, y_{\bar u}, p_{\bar u}, \bar u) z_v^2 + 2H_{uy}(x, y_{\bar u}, p_{\bar u}, \bar u) z_v v \Big]\, dx. \tag{10}$$

Remark 3.7. We mention that the quadratic form $\Lambda : L^2(\Omega) \to \mathbb{R}$ is the second variation of the objective functional $\mathcal{J} : \mathcal{U} \to \mathbb{R}$ at $\bar u$. By Proposition 3.5, we also have the following representation

$$\Lambda(v) = \int_\Omega \pi_v v\, dx \quad \forall v \in L^2(\Omega).$$

We close this section with a result concerning the quadratic form (10).

Proposition 3.8. *Let $\{v_k\}_{k=1}^\infty \subset L^2(\Omega)$ and $v \in L^2(\Omega)$. If $v_k \rightharpoonup v$ weakly in $L^2(\Omega)$, then $\Lambda(v_k) \to \Lambda(v)$.*

Proof. By Proposition 3.6, $\pi_{v_k} \to \pi_v$ in $L^\infty(\Omega)$, therefore

$$\Lambda(v_k) = \int_\Omega (\pi_{v_k} - \pi_v) v_k\, dx + \int_\Omega \pi_v v_k\, dx \to \int_\Omega \pi_v v\, dx.$$

□

4 Stability

In this section, we study the stability of the optimal solution of problem (1)–(2) with respect to perturbations. As usual in optimization, the stability of the solution is derived from stability of the system of necessary optimality conditions. The investigated stability property of the latter is the so-called strong metric Hölder subregularity (SMHSr), see e.g., [16, Section 3I] or [11, Section 4]. After introducing the assumptions we study the SMHSr property of the variational inequality (9). Then the result is used to obtain this property for the whole system of necessary optimality conditions

4.1 The main assumption

We begin the section recalling that $\bar u \in \mathcal{U}$ is a local minimizer of problem (1)–(2), and the definition of the quadratic form $\Lambda : L^2(\Omega) \to \mathbb{R}$ in (10).

Assumption 2. *There exist positive numbers α_0, γ_0 and $k^* \in [1, 4/n)$ such that*

$$\int_\Omega \sigma_{\bar u}(u - \bar u)\, dx + \Lambda(u - \bar u) \ge \gamma_0 |u - \bar u|_{L^1(\Omega)}^{k^*+1}, \tag{11}$$

for all $u \in \mathcal{U}$ with $|u - \bar u|_{L^1(\Omega)} \le \alpha_0$.

Assumption 2 resembles the well-known L^2-coercivity condition in optimal control, with two substantial differences: (i) the left-hand side of (11) involves a linear term (not only the quadratic form in the L^2-coercivity condition); (ii) the L^1-norm appears in the right-hand side of (11). Assumption 2 in the particular case $k^* = 1$ has been used before in the literature on optimal control problems constrained by ordinary differential equations, see [28, Assumption A2'] or [29, Assumption A2]. A similar assumption was used in [14, Assumption 2]. We first point out that if $\bar u$ satisfies Assumption 2, then it must be bang-bang. A control $u \in \mathcal{U}$ is bang-bang if $u(x) \in \{b_1(x), b_2(x)\}$ for a.e. x in Ω. The proof of this result follows the arguments given in the proof of [10, Theorem 2.1].

Proposition 4.1. *If $\bar{u} \in \mathcal{U}$ satisfies Assumption 2, then $\bar{u}$ is bang-bang.*

Proof. Let α_0 and γ_0 be the positive numbers in Assumption 2. Suppose that there exists $\varepsilon > 0$ and a measurable set $E \subset \Omega$ of positive measure such that

$$\bar{u}(x) \in [b_1(x) + \varepsilon, b_2(x) - \varepsilon] \quad \text{for a.e. } x \in E.$$

Define $\varepsilon^* := \min\{\alpha_0(\operatorname{meas} E)^{-1}, \varepsilon\}$. Let $\{v_m\}_{m=1}^{\infty} \subset L^2(\Omega)$ be a sequence converging to zero weakly in $L^2(\Omega)$ such that for each $m \in \mathbb{N}$, $v_m(x) \in \{-\varepsilon^*, \varepsilon^*\}$ for a.e. $x \in \Omega$. For each $m \in \mathbb{N}$, define

$$u_m(x) := \begin{cases} \bar{u}(x) & if \quad x \notin E \\ \bar{u}(x) + v_m(x) & if \quad x \in E. \end{cases}$$

Clearly, for each $m \in \mathbb{N}$, u_m belongs to $\mathcal{U}$ and

$$|u_m - \bar{u}|_{L^1(\Omega)} = \varepsilon^* \operatorname{meas} E.$$

Hence, by Assumption 2

$$\int_\Omega \sigma_{\bar{u}}(u_m - \bar{u})\, dx + \Lambda(u_m - \bar{u}) \geq \gamma_0 \Big(\varepsilon^* \operatorname{meas} E\Big)^{k^*+1} \tag{12}$$

for all $m \in \mathbb{N}$. Since $u_m \rightharpoonup \bar{u}$ weakly in $L^2(\Omega)$, we have by Proposition 3.8 that the left hand side of (12) converges to 0; a contradiction. □

Proposition 4.1 makes the following lemma relevant.

Lemma 4.2. *Let $u \in \mathcal{U}$ be bang-bang, and $\{u_k\}_{k=1}^{\infty} \subset \mathcal{U}$ be a sequence. If $u_k \rightharpoonup u$ weakly in $L^1(\Omega)$, then $u_k \to u$ in $L^1(\Omega)$.*

Proof. Let $\Omega_i := \{x \in \Omega : u(x) = b_i(x)\}$, $i = 1, 2$. Let $\chi_{\Omega_i} : \Omega \to \{0, 1\}$ denote the characteristic function of the set Ω_i, $i = 1, 2$. Now, by definition of weak convergence

$$\int_\Omega |u_k - u|\, dx = \int_\Omega \chi_{\Omega_1}(u_n - \bar{u})\, dx - \int_\Omega \chi_{\Omega_2}(u_n - \bar{u})\, dx \to 0.$$

□

The next proposition shows that the switching mapping satisfies a growth condition. The proof consists of two steps. The first one is to show that Assumption 2 implies this growth condition for the linearization of the switching mapping. The second step is to adequately use the linearization as an approximation of the switching mapping.

Proposition 4.3. *Let Assumption 2 be fulfilled. Then there exist positive numbers α and γ such that*

$$\int_\Omega \sigma_u(u - \bar{u})\, dx \geq \gamma |u - \bar{u}|_{L^1(\Omega)}^{k^*+1}$$

for all $u \in \mathcal{U}$ with $|u - \bar{u}|_{L^1(\Omega)} \leq \alpha$.

Proof. Let α_0, γ_0 and k^* be the positive numbers in Assumption 2. Fix $r \in (n/2, 2/k^*)$. Using Proposition 3.4, a constant $c > 0$ can be found such that

$$|\sigma_u - \sigma_{\bar{u}} - \pi_{u-\bar{u}}|_{L^\infty(\Omega)} \leq c|u - \bar{u}|_{L^1(\Omega)}^{2/r} \qquad \forall u \in \mathcal{U}. \tag{13}$$

From Proposition 3.5 and Assumption 2, we have

$$\int_\Omega \Big[\sigma_{\bar{u}} + \pi_{u-\bar{u}}\Big](u - \bar{u})\, dx = \int_\Omega \sigma_{\bar{u}}(u - \bar{u})\, dx + \Lambda(u - \bar{u}) \geq \gamma_0 |u - \bar{u}|_{L^1(\Omega)}^{k^*+1} \tag{14}$$

for all $u \in \mathcal{U}$ with $|u-\bar u|_{L^1(\Omega)} \le \alpha_0$. Define $\gamma := \gamma_0/2$ and

$$\alpha := \min\left\{\alpha_0, \gamma^{\frac{r}{2-k^* r}} c^{-\frac{r}{2-k^* r}}\right\}.$$

Then, by (13)

$$|\sigma_u - \sigma_{\bar u} - \pi_{u-\bar u}|_{L^\infty(\Omega)} \le c|u-\bar u|_{L^1(\Omega)}^{\frac{2}{r}} = c|u-\bar u|_{L^1(\Omega)}^{\frac{2}{r}-k^*}|u-\bar u|_{L^1(\Omega)}^{k^*} \le \gamma |u-\bar u|_{L^1(\Omega)}^{k^*} \tag{15}$$

for all $u \in \mathcal{U}$ with $|u-\bar u|_{L^1(\Omega)} \le \alpha$. We have for all $u \in \mathcal{U}$

$$\int_\Omega \sigma_u(u-\bar u)\,dx = \int_\Omega \Big[\sigma_{\bar u} + \pi_{u-\bar u}\Big](u-\bar u)\,dx + \int_\Omega \Big[\sigma_u - \sigma_{\bar u} - \pi_{u-\bar u}\Big](u-\bar u)\,dx.$$

Consequently, by (14) and (15),

$$\begin{aligned}\int_\Omega \sigma_u(u-\bar u)\,dx &\ge \gamma_0|u-\bar u|_{L^1(\Omega)}^{k^*+1} - |\sigma_u - \sigma_{\bar u} - \pi_{u-\bar u}|_{L^\infty(\Omega)}|u-\bar u|_{L^1(\Omega)} \\ &= (\gamma_0-\gamma)|u-\bar u|_{L^1(\Omega)}^{k^*+1} = \gamma|u-\bar u|_{L^1(\Omega)}^{k^*+1}\end{aligned}$$

for all $u \in \mathcal{U}$ with $|u-\bar u|_{L^1(\Omega)} \le \alpha$. □

4.2 Some existence and stability results

We now pass to some preparatory lemmas concerning the existence of solutions of inclusions (also called generalized equations, see [36]) related to the first order necessary condition of problem (1)–(2). Given $r \in [1,\infty]$, we denote by $\mathbb{B}_{L^r}(c;\alpha)$ the closed ball in $L^r(\Omega)$ with center $c \in L^r(\Omega)$ and radius $\alpha > 0$.

The variational inequality (9) can be written as the inclusion

$$0 \in \sigma_u + N_{\mathcal{U}}(u),$$

where the normal cone at u to the set $\mathcal{U}$ is given by

$$N_{\mathcal{U}}(u) = \left\{\sigma \in L^\infty(\Omega) : \int_\Omega \sigma(w-u)\,dx \le 0 \quad \forall w \in \mathcal{U}\right\}$$

Lemma 4.4. *For all $\rho \in L^\infty(\Omega)$ and $\varepsilon > 0$ there exists $u \in \mathcal{U} \cap \mathbb{B}_{L^1}(\bar u;\varepsilon)$ satisfying*

$$\rho \in \sigma_u + N_{\mathcal{U}\cap\mathbb{B}_{L^1}(\bar u;\varepsilon)}(u).$$

Proof. Let $\rho \in L^\infty(\Omega)$ and $\varepsilon > 0$. Consider the functional $\mathcal{J}_\rho : \mathcal{U} \to \mathbb{R}$

$$\mathcal{J}_\rho(u) := \int_\Omega \big[g(y_u,u) - \rho u\big]\,dx = \mathcal{J}(u) - \int_\Omega \rho u\,dx.$$

The functional $\mathcal{J}_\rho$ has at least one global minimizer $u_\rho \in \mathcal{U}$ since $\mathcal{U} \cap \mathbb{B}_{L_1}(\bar u;\varepsilon)$ is a weakly sequentially compact subset of $L^2(\Omega)$ and $\mathcal{J}_\rho$ is weakly sequentially continuous. By the Pontryagin principle,

$$\int_\Omega \big[\sigma_{u_\rho} - \rho\big](u-u_\rho)\,dx \ge 0 \quad \forall u \in \mathcal{U}.$$

We have then that u_ρ satisfies $\rho \in \sigma_{u_\rho} + N_{\mathcal{U}\cap\mathbb{B}_{L^1}(\bar u;\varepsilon)}(u_\rho)$. □

Lemma 4.5. *Let $\mathcal{V}_1$ and $\mathcal{V}_2$ be closed and convex subsets of $L^1(\Omega)$ such that $\mathcal{V}_1 \cap \operatorname{int}\mathcal{V}_2 \ne \emptyset$. Then*

$$N_{\mathcal{V}_1\cap\mathcal{V}_2}(u) = N_{\mathcal{V}_1}(u) + N_{\mathcal{V}_2}(u) \tag{16}$$

for all $u \in \mathcal{V}_1 \cap \mathcal{V}_2$.

Proof. Given a set $\mathcal{W} \subset L^1(\Omega)$, let $s_{\mathcal{W}} : L^\infty(\Omega) \to \mathbb{R} \cup \{\infty\}$ denote the support function to $\mathcal{W}$, that is

$$s_{\mathcal{W}}(h) := \sup_{w \in \mathcal{W}} \int_\Omega hw \, dx.$$

By [3, Proposition 3.1], the set $\operatorname{Epi} s_{\mathcal{V}_1} + \operatorname{Epi} s_{\mathcal{V}_2}$ is weak* closed in $L^\infty(\Omega)$. Then the representation (16) holds according to [3, Theorem 3.1]. □

We can now prove existence of solutions of the inclusion $\rho \in \sigma_u + N_{\mathcal{U}}(u)$ that are close (in the L^1-norm) to $\bar{u}$ whenever ρ is close to zero (in the norm L^∞-norm). The proof follows the arguments in [13, p. 1127].

Lemma 4.6. *Let Assumption 2 hold. For each $\varepsilon > 0$ there exists $\delta > 0$ such that for each $\rho \in \mathbb{B}_{L^\infty}(0;\delta)$ there exists $u \in \mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u};\varepsilon)$ satisfying $\rho \in \sigma_u + N_{\mathcal{U}}(u)$.*

Proof. Let α and γ be the numbers in Proposition 4.3. Define $\varepsilon_0 := \min\{\varepsilon, \alpha\}$ and $\delta := \varepsilon_0^{k^*}\gamma/2$. Let $\rho \in L^\infty(\Omega)$ with $|\rho|_{L^\infty(\Omega)} \leq \delta$. By Lemma 4.4, there exists $u \in \mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u};\varepsilon_0)$ such that

$$\rho \in \sigma_u + N_{\mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u};\varepsilon_0)}(u).$$

Since trivially $\bar{u} \in \mathcal{U} \cap \operatorname{int} \mathbb{B}_{L^1}(\bar{u}, \varepsilon_0)$, by Lemma 4.5 we have

$$N_{\mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u};\varepsilon_0)}(u) = N_{\mathcal{U}}(u) + N_{\mathbb{B}_{L^1}(\bar{u};\varepsilon_0)}(u). \tag{17}$$

Thus there exists $\nu \in N_{\mathbb{B}_{L^1}(\bar{u};\varepsilon_0)}(u)$ such that

$$\rho - \sigma_u - \nu \in N_{\mathcal{U}}(u).$$

By definition of the normal cone,

$$0 \geq \int_\Omega (\rho - \sigma_u)(\bar{u} - u) \, dx - \int_\Omega \nu(\bar{u} - u) \, dx. \tag{18}$$

As $\bar{u} \in \mathbb{B}_{L^1}(\bar{u};\varepsilon_0)$ and $\nu \in N_{\mathbb{B}_{L^1}(\bar{u};\varepsilon_0)}(u)$, we have

$$\int_\Omega \nu(\bar{u} - u) \, dx \leq 0.$$

Consequently, by (18) and Proposition 4.3

$$0 \geq \int_\Omega (\rho - \sigma_u)(\bar{u} - u) \, dx \geq -|\rho|_{L^\infty(\Omega)} |u - \bar{u}|_{L^1(\Omega)} + \gamma |u - \bar{u}|_{L^1(\Omega)}^{k^*+1},$$

which implies

$$|u - \bar{u}|_{L^1(\Omega)} \leq \gamma^{-\frac{1}{k^*}} |\rho|_{L^\infty(\Omega)}^{\frac{1}{k^*}} \leq 2^{-\frac{1}{k^*}} \varepsilon_0 < \varepsilon_0.$$

As $u \in \operatorname{int} \mathbb{B}_{L^1}(\bar{u};\varepsilon_0)$, we have $N_{\mathbb{B}_{L^1}(\bar{u};\varepsilon_0)}(u) = \{0\}$. Thus by (17),

$$\rho \in \sigma_u + N_{\mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u};\varepsilon_0)}(u) = \sigma_u + N_{\mathcal{U}}(u). \tag{19}$$

□

The following lemma shows how Proposition 4.3 (and consequently Assumption 2) is related to Hölder-stability.

Lemma 4.7. *Let Assumption 2 hold. There exist positive numbers α and c such that*

$$|u - \bar{u}|_{L^1(\Omega)} \leq c |\rho|_{L^\infty(\Omega)}^{\frac{1}{k^*}} \tag{20}$$

for all $\rho \in L^\infty(\Omega)$ and $u \in \mathbb{B}_{L^1}(\bar{u};\alpha)$ satisfying $\rho \in \sigma_u + N_{\mathcal{U}}(u)$.

Proof. Let α and γ be the positive numbers in Proposition 4.3. Since $\rho - \sigma_u \in N_{\mathcal{U}}(u)$, we have

$$\int_\Omega (\rho - \sigma_u)(\bar{u} - u)\, dx \le 0.$$

By Proposition 4.3,

$$\begin{aligned} 0 &\ge \int_\Omega (\rho - \sigma_u)(\bar{u} - u)\, dx = \int_\Omega \sigma_u (u - \bar{u})\, dx + \int_\Omega \rho(\bar{u} - u)\, dx \\ &\ge \gamma \left(\int_\Omega |u - \bar{u}|\, dx \right)^{k^*+1} - |\rho|_{L^\infty(\Omega)} \int_\Omega |u - \bar{u}|\, dx. \end{aligned}$$

Hence

$$\int_\Omega |u - \bar{u}|\, dx \le \left(\frac{1}{\gamma} |\rho|_{L^\infty(\Omega)} \right)^{1/k^*} = \gamma^{-\frac{1}{k^*}} |\rho|_{L^\infty(\Omega)}^{\frac{1}{k^*}}.$$

The result follows defining $c = \gamma^{-\frac{1}{k^*}}$. □

Lemma 4.7 requires that the controls are close (in the L^1-norm) a priori for the inequality (20) to hold. This assumption can be removed if the solution of the inclusion $0 \in \sigma_u + N_{\mathcal{U}}(u)$ is unique.

Lemma 4.8. *Let Assumption 2 hold, and suppose additionally that $\bar{u} \in \mathcal{U}$ is the unique solution of $0 \in \sigma_u + N_{\mathcal{U}}(u)$. There exist positive numbers δ and c such that*

$$|u - \bar{u}|_{L^1(\Omega)} \le c |\rho|_{L^\infty(\Omega)}^{\frac{1}{k^*}}.$$

for all $\rho \in \mathbb{B}_{L^\infty}(0;\delta)$ and $u \in \mathcal{U}$ satisfying $\rho \in \sigma_u + N_{\mathcal{U}}(u)$.

Proof. Let α and c be the positive numbers in Lemma 4.7. First we prove that there exists $\delta > 0$ such that if $u \in \mathcal{U}$ and $\rho \in L^\infty(\Omega)$ satisfy $\rho \in \sigma_u + N_{\mathcal{U}}(u)$ and $|\rho|_{L^\infty(\Omega)} \le \delta$, then $u \in \mathbb{B}_{L^1}(\bar{u};\alpha)$. Suppose not, then there exist sequences $\{\rho_k\}_{k=1}^\infty \subset L^\infty(\Omega)$ and $\{u_k\}_{k=1}^\infty \subset \mathcal{U}$ such that $\rho_k \in \sigma_{u_k} + N_{\mathcal{U}}(u_k)$ and $|u_k - \bar{u}|_{L^1(\Omega)} > \alpha$. Since $\mathcal{U}$ is weakly sequentially compact in $L^2(\Omega)$, there exists a subsequence of $\{u_k\}_{k=1}^\infty$, denoted in the same way, and $u^* \in \mathcal{U}$ such that $u_k \rightharpoonup u^*$ weakly in $L^2(\Omega)$. Using Proposition 2.12, one can see that $\rho_k - \sigma_{u_k} \to \sigma_{u^*}$ in $L^\infty(\Omega)$. Consequently, as $\rho_k \in \sigma_{u_k} + N_{\mathcal{U}}(u_k)$ for all $n \in \mathbb{N}$, we obtain $0 \in \sigma_{u^*} + N_{\mathcal{U}}(u^*)$. Then, by assumption, $u^* = \bar{u}$, so u^* is bang-bang. By Lemma 4.2, we have $u_k \to u^*$ in $L^1(\Omega)$; a contradiction. The result follows from Lemma 4.7. □

4.3 Strong metric subregularity

Let us begin considering the following system representing the necessary optimality conditions (Pontryagin principle) for problem (1)–(2):

$$\begin{cases} 0 &= \mathcal{L}y - f(\cdot, y, u), \\ 0 &= \mathcal{L}p - H_y(\cdot, y, p, u), \\ 0 &\in H_u(\cdot, y, p) + N_{\mathcal{U}}(u), \end{cases} \tag{21}$$

If $u \in \mathcal{U}$ is a local solution of problem (1)–(2), then the triple (y_u, p_u, u) is a solution of (21). Therefore, the mapping that defines the right-hand side is referred to as the *optimality mapping*. In order to give a strict definition and recast system (21) in a functional frame, we introduce the metric spaces

$$\mathcal{Y} := D(\mathcal{L}) \times D(\mathcal{L}) \times \mathcal{U} \quad \text{and} \quad \mathcal{Z} := L^2(\Omega) \times L^2(\Omega) \times L^\infty(\Omega),$$

endowed with the following metrics. For $\psi_i = (y_i, p_i, u_i) \in \mathcal{Y}$ and $\zeta_i = (\xi_i, \eta_i, \rho_i) \in \mathcal{Z}$, $i \in \{1, 2\}$,

$$\begin{aligned} d_{\mathcal{Y}}(\psi_1, \psi_2) &:= |y_1 - y_2|_{L^2(\Omega)} + |p_1 - p_2|_{L^2(\Omega)} + |u_1 - u_2|_{L^1(\Omega)}, \\ d_{\mathcal{Z}}(\zeta_1, \zeta_2) &:= |\xi_1 - \xi_2|_{L^2(\Omega)} + |\eta_1 - \eta_2|_{L^2(\Omega)} + |\rho_1 - \rho_2|_{L^\infty(\Omega)}. \end{aligned}$$

Both metrics are shift-invariant. We denote by $\mathbb{B}_{\mathcal{Y}}(\psi;\alpha)$ the closed ball in $\mathcal{Y}$, centered at ψ and with radius α. The notation for the ball $\mathbb{B}_{\mathcal{Z}}(\zeta;\alpha)$ is identical. Then the optimality mapping is defined as the set-valued mapping $\Phi:\mathcal{Y}\twoheadrightarrow\mathcal{Z}$ given by

$$\Phi(y,p,u)=\begin{pmatrix}\mathcal{L}y-f(\cdot,y,u)\\ \mathcal{L}p-H_y(\cdot,y,p,u)\\ H_u(\cdot,y,p,u)+N_{\mathcal{U}}(u)\end{pmatrix}. \tag{22}$$

Then the optimality system (21) can be recast as the inclusion

$$0\in\Phi(y,p,u). \tag{23}$$

Our purpose is to study the stability of system (21), or equivalently of inclusion (23), with respect to perturbations in the right-hand side. From now on, we denote $\bar\psi:=(\bar y,\bar p,\bar u)=(y_{\bar u},p_{\bar u},\bar u)$, where $\bar u$ is the fixed local solution of problem (1)–(2).

Definition 4.9. The optimality mapping $\Phi:\mathcal{Y}\twoheadrightarrow\mathcal{Z}$ is called strongly Hölder subregular with exponent $\lambda>0$ at $(\bar\psi,0)$ if there exist positive numbers α_1,α_2 and κ such that

$$d_{\mathcal{Y}}(\psi,\bar\psi)\le\kappa d_{\mathcal{Z}}(\zeta,0)^{\lambda} \tag{24}$$

for all $\psi\in\mathbb{B}_{\mathcal{Y}}(\bar\psi;\alpha_1)$ and $\zeta\in\mathbb{B}_{\mathcal{Z}}(0;\alpha_2)$ satisfying $\zeta\in\Phi(\psi)$.

More explicitly, the inequality (24) reads as

$$|y-y_{\bar u}|_{L^2(\Omega)}+|p-p_{\bar u}|_{L^2(\Omega)}+|u-\bar u|_{L^1(\Omega)}\le\kappa\Big(|\xi|_{L^2(\Omega)}+|\eta|_{L^2(\Omega)}+|\rho|_{L^\infty(\Omega)}\Big)^{\lambda}. \tag{25}$$

Hence, if the optimality mapping is strongly Hölder subregular, all solutions of the system

$$\begin{cases}\xi &=& \mathcal{L}y-f(\cdot,y,u),\\ \eta &=& \mathcal{L}p-H_y(\cdot,y,p,u),\\ \rho &\in& H_u(\cdot,y,p)+N_{\mathcal{U}}(u).\end{cases} \tag{26}$$

that are near $(y_{\bar u},p_{\bar u},\bar u)$ satisfy the Hölder estimate (25) with respect to the perturbations $\zeta=(\xi,\eta,\rho)$, provided they are small enough.

Remark 4.10. If Φ is strongly Hölder subregular at $(\bar\psi,0)$, then from (24) applied with $\zeta=0$ we obtain that $\bar\psi$ is the unique solution of (23) in $B_{\mathcal{Y}}(\bar\psi;\alpha_1)$, hence $\bar u$ is the unique local solution of problem (1)–(2) in this ball. In particular, $\bar u$ is a strict local minimizer.

We are now ready to state our main result.

Theorem 4.11. *Let Assumption 2 hold. Then the optimality mapping Φ is strongly Hölder subregular at $(\bar\psi,0)$ with exponent $\lambda=1/k^*$.*

Proof. Let α and c be the positive numbers in Lemma 4.7. Let $\zeta=(\xi,\eta,\rho)\in B_{\mathcal{Z}}(0;1)$ and $\psi=(y,p,u)\in\mathbb{B}_{\mathcal{Y}}(\bar\psi;\alpha)$ such that $\zeta\in\Phi(\psi)$. By a standard argument, there exists $c_1>0$ (independent of ψ and ζ) such that

$$|y-y_u|_{L^\infty(\Omega)}+|p-p_u|_{L^\infty(\Omega)}\le c_1\Big(|\xi|_{L^2(\Omega)}+|\eta|_{L^2(\Omega)}\Big). \tag{27}$$

Since H_u is locally Lipschitz uniformly in the first variable, and the sets $\{y_u:u\in\mathcal{U}\}$, $\{p_u:u\in\mathcal{U}\}$ are bounded in $C(\bar\Omega)$, there exists $c_2>0$ (independent of ψ) such that

$$|H_u(\cdot,y,p)-H_u(\cdot,y_u,p_u)|_{L^\infty(\Omega)}\le c_2\Big(|y-y_u|_{L^\infty(\Omega)}+|p-p_u|_{L^\infty(\Omega)}\Big) \tag{28}$$

Define $\nu:=\rho+H_u(\cdot,y_u,p_u)-H_u(\cdot,y,p)$. By (27) and (28), there exists $c_3>0$ (independent of ψ and ζ) such that

$$|\nu|_{L^\infty(\Omega)}\le c_3\Big(|\xi|_{L^2(\Omega)}+|\eta|_{L^2(\Omega)}+|\rho|_{L^\infty(\Omega)}\Big)=c_3|\zeta|_{\mathcal{Z}}.$$

As $\rho \in H_u(\cdot, y, p) + N_{\mathcal{U}}(u)$, we have $\nu \in H_u(\cdot, y_u, p_u) + N_{\mathcal{U}}(u)$. Then by Lemma 4.7,

$$|u - \bar{u}|_{L^1(\Omega)} \le c|\nu|_{L^\infty(\Omega)}^{\frac{1}{k^*}} \le c c_3^{\frac{1}{k^*}} |\zeta|_{\mathcal{Z}}^{\frac{1}{k^*}} := c_4 |\zeta|_{\mathcal{Z}}^{\frac{1}{k^*}}. \tag{29}$$

Now, by Proposition 2.11, there exists $c_5 > 0$ (independent of ψ) such that $|y_u - y_{\bar{u}}|_{L^2(\Omega)} \le c_5 |u - \bar{u}|_{L^1(\Omega)}$. Consequently, by (29)

$$\begin{aligned} |y - y_{\bar{u}}|_{L^2(\Omega)} &\le |y - y_u|_{L^2(\Omega)} + |y_u - y_{\bar{u}}|_{L^2(\Omega)} \\ &\le c_1 \text{meas}\, \Omega^{\frac{1}{2}} \Big(|\xi|_{L^2(\Omega)} + |\eta|_{L^2(\Omega)} \Big) + c_5 |u - \bar{u}|_{L^1(\Omega)} \\ &\le (c_1 \text{meas}\, \Omega^{\frac{1}{2}} + c_5 c_4) |\zeta|_{\mathcal{Z}}^{\frac{1}{k^*}} =: c_6 |\zeta|_{\mathcal{Z}}^{\frac{1}{k^*}}. \end{aligned}$$

Analogously, there exists $c_7 > 0$ (independent of ψ and ζ) such that

$$|p - p_u|_{L^2(\Omega)} \le c_7 |\zeta|_{\mathcal{Z}}^{\frac{1}{k^*}}.$$

Putting all together,

$$|y - y_{\bar{u}}|_{L^2(\Omega)} + |p - p_u|_{L^2(\Omega)} + |u - \bar{u}|_{L^1(\Omega)} \le (c_4 + c_6 + c_7) |\zeta|_{\mathcal{Z}}^{\frac{1}{k^*}}.$$

Finally, let $\alpha_1 := \alpha$, $\alpha_2 := 1$ and $\kappa := c_4 + c_6 + c_7$. Since the constants c_4, c_6 and c_7 are independent of ψ and ζ, so is κ. Thus we have (24) for all $\psi \in \mathbb{B}_{\mathcal{Y}}(\bar{\psi}; \alpha_1)$ and $\zeta \in \mathbb{B}_{\mathcal{Z}}(0; \alpha_2)$ satisfying $\zeta \in \Phi(\psi)$. □

The strong subregularity property defined above does not require existence of solutions of the perturbed inclusion (26) in a neighborhood of the reference solution $\bar{\psi}$. The next theorem answers the existence question.

Theorem 4.12. *Let Assumption 2 hold. For each $\varepsilon > 0$ there exists $\delta > 0$ such that for every $\zeta \in \mathbb{B}_{\mathcal{Z}}(0; \delta)$ there exists $\psi \in \mathbb{B}_{\mathcal{Y}}(\bar{\psi}; \varepsilon)$ satisfying the inclusion $\zeta \in \Phi(\psi)$.*

Proof. For each $u \in \mathcal{U}$ and $\zeta = (\xi, \eta, \rho) \in \mathcal{Z}$, define $\nu_{u,\zeta} := \rho + H_u(\cdot, y_u, p_u) - H_u(\cdot, y_{u,\zeta}, p_{u,\zeta})$, where $y_{u,\zeta}$ and $p_{u,\zeta}$ are the unique solutions of

$$\begin{cases} \mathcal{L} y &= f(\cdot, y, u) + \xi, \\ \mathcal{L} p &= H_y(\cdot, y, p, u) + \eta. \end{cases} \tag{30}$$

By a standard argument, one can find positive numbers c_1 and c_2 such that

$$|y_{u,\zeta} - y_u|_{L^2(\Omega)} + |p_{u,\zeta} - p_u|_{L^2(\Omega)} \le c_1 \Big(|\xi|_{L^2(\Omega)} + |\eta|_{L^2(\Omega)} \Big), \tag{31}$$

and $|\nu_{u,\zeta}|_{L^\infty(\Omega)} \le c_2 |\zeta|_{\mathcal{Z}}$ for all $u \in \mathcal{U}$ and $\zeta \in \mathcal{Z}$. Let $\varepsilon > 0$ be arbitrary. By Lemma 4.6, the exists $\delta_0 > 0$ such that for each $\nu \in \mathbb{B}_{L^\infty}(0; \delta_0)$ there exists $u \in \mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u}; \varepsilon/2)$ satisfying $\nu \in \sigma_u + N_{\mathcal{U}}(u)$. Define $\delta := \min\{c_2^{-1}\delta_0, (2c_1)^{-1}\varepsilon\}$ and let $\zeta^* \in \mathbb{B}_{\mathcal{Z}}(0; \delta)$ be arbitrary; we will prove that there exists $u^* \in \mathcal{U} \cap \mathbb{B}_{L^1}(\bar{u}; \varepsilon/2)$ such that $\nu_{u^*,\zeta^*} \in \sigma_{u^*} + N_{\mathcal{U}}(u^*)$. First, observe that

$$|\nu_{u,\zeta^*}|_{L^\infty(\Omega)} \le c_2 |\zeta^*|_{\mathcal{Z}} \le \delta_0 \quad \forall u \in \mathcal{U}.$$

Therefore, by Lemma 4.6, we can inductively define a sequence $\{u_k\}_{k=1}^\infty \subset \mathcal{U}$ such that $\nu_{u_k,\zeta^*} \in \sigma_{u_{k+1}} + N_{\mathcal{U}}(u_{k+1})$ and $|u_k - \bar{u}|_{L^1(\Omega)} \le \varepsilon/2$ for all $k \in \mathbb{N}$. Since $\mathcal{U}$ is weakly compact in $L^2(\Omega)$, we may assume that $u_k \rightharpoonup u^*$ weakly in $L^2(\Omega)$ for some $u^* \in \mathcal{U}$. Weak convergence in $L^2(\Omega)$ implies weak convergence in $L^1(\Omega)$ and $\mathbb{B}_{L^1}(\bar{u}; \varepsilon/2)$ is weakly sequentially closed in $L^1(\Omega)$, therefore $u^* \in \mathbb{B}_{L^1}(\bar{u}; \varepsilon/2)$. Using Proposition 2.12, one can see that $\nu_{u_k,\zeta^*} - \sigma_{u_{k+1}} \to \nu_{u^*,\zeta^*} - \sigma_{u^*}$ in $L^\infty(\Omega)$, and consequently that $\nu_{u^*,\zeta^*} \in \sigma_{u^*} + N_{\mathcal{U}}(u^*)$. We conclude then that $\zeta^* \in \Phi(\psi^*)$, where $\psi^* := (y_{u^*,\zeta^*}, p_{u^*,\zeta^*}, u^*)$. Finally, by definition of δ and (31)

$$|\psi^* - \bar{\psi}|_{\mathcal{Y}} \le c_1 |\zeta|_{\mathcal{Z}} + \varepsilon/2 \le \varepsilon.$$

Thus, $\zeta^* \in \Phi(\psi^*)$ and $\psi^* \in \mathbb{B}_{\mathcal{Y}}(\bar{\psi}; \varepsilon)$, which completes the proof. □

The next theorem claims that *all* solutions of the perturbed optimality system (26) are arbitrarily close to the solution of the unperturbed optimality system, provided that the solution of the latter is globally unique, Assumption 2 holds, and the perturbation is sufficiently small.

Theorem 4.13. *Let Assumption 2 hold and suppose additionally that $\bar\psi$ is the unique element of $\mathcal{Y}$ that satisfies $0 \in \Phi(\bar\psi)$. For each $\varepsilon > 0$ there exists $\delta > 0$ such that if $\zeta \in \mathbb{B}_{\mathcal{Z}}(0;\delta)$ and $\psi \in \mathcal{Y}$ satisfy $\zeta \in \Phi(\psi)$, then $\psi \in \mathbb{B}_{\mathcal{Y}}(\bar\psi;\varepsilon)$.*

Proof. Let δ_0 and c_0 be the positive numbers in Lemma 4.8. Let $\zeta = (\xi,\eta,\rho) \in \mathcal{Z}$ and $\psi = (y,p,u) \in \mathcal{Y}$ be such that $\zeta \in \Phi(\psi)$. Define $\nu := \rho + H_u(\cdot, y_u, p_u) - H_u(\cdot, y, p)$. Arguing as in the proof of Theorem 4.11, we can find positive numbers c_1 and c_2 (independent of ψ and ζ) such that $|\nu|_{L^\infty(\Omega)} \le c_1|\zeta|_{\mathcal{Z}}$ and

$$|y - y_{\bar u}|_{L^2(\Omega)} + |p - p_{\bar u}|_{L^2(\Omega)} \le c_2\Big(|\zeta|_{\mathcal{Z}} + |u - \bar u|_{L^1(\Omega)}\Big).$$

Let $\delta := \min\{c_1^{-1}\delta_0, (2c_0c_2)^{-k^*}c_1^{-1}\varepsilon^{k^*}, (2c_2)^{-1}\varepsilon\}$ and suppose that $\zeta \in \mathbb{B}_{\mathcal{Z}}(0;\delta)$. As $\rho \in H_u(\cdot,y,p) + N_{\mathcal{U}}(u)$, we have $\nu \in H_u(\cdot, y_u, p_u) + N_{\mathcal{U}}(u)$. By Lemma 4.8,

$$|u - \bar u|_{L^1(\Omega)} \le c_0|\nu|_{L^\infty(\Omega)}^{\frac{1}{k^*}} \le c_0 c_1^{\frac{1}{k^*}}|\zeta|_{\mathcal{Z}}^{\frac{1}{k^*}} \le c_2^{-1}\varepsilon/2.$$

Thus,

$$|y - y_{\bar u}|_{L^2(\Omega)} + |p - p_{\bar u}|_{L^2(\Omega)} + |u - \bar u|_{L^1(\Omega)} \le c_2\Big(\delta + c_2^{-1}\varepsilon/2\Big) \le \varepsilon.$$

□

5 Nonlinear Perturbations

In this section we apply the subregularity results in Section 4 for studying the effect of certain nonlinear perturbations on the optimal solution. We consider the following family of problems

$$\min_{u\in\mathcal{U}}\left\{\int_\Omega \Big[g(x,y,u) + \eta(x,y,u)\Big]\,dx\right\}, \tag{32}$$

subject to

$$\begin{cases} -\operatorname{div}\big(A(x)\nabla y\big) + d(x,y) + \xi(x,y) = \beta(x)u & \text{in} \quad \Omega \\ A(x)\nabla y\cdot\nu + b(x)y = 0 & \text{on} \quad \partial\Omega. \end{cases} \tag{33}$$

In order to specify the perturbations under consideration and their topology, we begin the section recalling some elementary notions of functional analysis.

As usual, $C(\mathbb{R}^s)$ denotes the space of all continuous functions $\omega : \mathbb{R}^s \to \mathbb{R}$. For each $m \in \mathbb{N}$, let K_m denote the closed ball in $\mathbb{R}^s$ centered at zero with radius m. Consider the metric on $C(\mathbb{R}^s)$ given by

$$d_C(\omega_1,\omega_2) := \sum_{m=1}^\infty \frac{1}{2^m}\frac{|\omega_1-\omega_2|_{L^\infty(K_m)}}{1+|\omega_1-\omega_2|_{L^\infty(K_m)}}.$$

This metric induces the compact-convergence topology on $C(\mathbb{R}^s)$. In this topology, a sequence $\{\omega_m\}_{m=1}^\infty \subset C(\mathbb{R}^s)$ converges to $\omega \in C(\mathbb{R}^s)$ if and only if $|\omega - \omega_m|_{L^\infty(K)} \to 0$ for every compact set $K \subset \mathbb{R}^s$. This topology is also known as the compact-open topology, see [26, Chapter 7].

Lemma 5.1. *For each compact set $K \subset \mathbb{R}^s$ there exists $m \in \mathbb{N}$ such that*

$$|\omega_1 - \omega_2|_{L^\infty(K)} \le 2^m d_C(\omega_1,\omega_2)$$

for all $\omega_1,\omega_2 \in C(\mathbb{R}^s)$ such that $d_C(\omega_1,\omega_2) \le 2^{-m}$.

5.1 The perturbations

We begin describing the space of perturbations appearing in equation (33). Let Υ_s be the set of all continuously differentiable functions $\xi : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ such that $d_y(x,y) + \xi_y(x,y) \geq 0$ for all $x \in \Omega$ and $y \in \mathbb{R}$. The set Υ_s does not constitute a linear space, but it allows to have well-defined states for each perturbation.

Proposition 5.2. *For each $u \in \mathcal{U}$ and $\xi \in \Upsilon_s$ there exists a unique function $y_u^\xi \in D(\mathcal{L})$ satisfying*

$$\mathcal{L}y_u^\xi + d(\cdot, y_u^\xi) + \xi(\cdot, y_u^\xi) = \beta(\cdot)u.$$

Moreover, there exist positive numbers M and δ such that $|y_u^\xi|_{L^\infty(\Omega)} \leq M$ for all $u \in \mathcal{U}$ and $\xi \in \Upsilon_s$ with $d_C(\xi, 0) \leq \delta$.

Proof. The existence follows from [39, Theorem 4.8]. Moreover, also from this theorem, there exists $c > 0$ such that

$$|y_u^\xi|_{L^\infty(\Omega)} \leq c\big|\beta(\cdot)u - d(\cdot,0) - \xi(\cdot,0)\big|_{L^\infty(\Omega)}$$

for all $u \in \mathcal{U}$ and $\xi \in \Upsilon_s$. Let $K := \bar{\Omega} \times \{0\}$, then by Lemma 5.1 there exists $m \in \mathbb{N}$ such that

$$\begin{aligned}
|y_u^\xi|_{L^\infty(\Omega)} &\leq c\Big(|\beta|_{L^\infty(\Omega)}|u|_{L^\infty(\Omega)} + |d(\cdot,0)|_{L^\infty(\Omega)} + |\xi|_{L^\infty(K)}\Big) \\
&\leq c\Big(|\beta|_{L^\infty(\Omega)} \sup_{u\in\mathcal{U}} |u|_{L^\infty(\Omega)} + |d(\cdot,0)|_{L^\infty(\Omega)} + 2^m d_C(\xi,0)\Big) \\
&\leq c\Big(|\beta|_{L^\infty(\Omega)} \sup_{u\in\mathcal{U}} |u|_{L^\infty(\Omega)} + |d(\cdot,0)|_{L^\infty(\Omega)} + 1\Big)
\end{aligned}$$

for all $u \in \mathcal{U}$ and $\xi \in \Upsilon_s$ with $d_C(\xi,0) \leq 2^{-m}$. The result follows defining $\delta := 2^{-m}$ and

$$M := c\Big(|\beta|_{L^\infty(\Omega)} \sup_{u\in\mathcal{U}} |u|_{L^\infty(\Omega)} + |d(\cdot,0)|_{L^\infty(\Omega)} + 1\Big).$$

□

We now proceed to describe the perturbations appearing in the cost functional (32). Consider the set Υ_c of all continuously differentiable functions $\eta : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that $\eta(x,y,\cdot)$ is convex for all $x \in \Omega$ and $y \in \mathbb{R}$. We have the following result concerning the adjoint variable of the perturbed problem. Its proof is similar to the one of Proposition 5.2.

Proposition 5.3. *For each $u \in \mathcal{U}$, $\xi \in \Upsilon_s$ and $\eta \in \Upsilon_c$ there exists a unique function $p_u^{\xi,\eta} \in D(\mathcal{L})$ satisfying*

$$\mathcal{L}p_u^{\xi,\eta} + \big[d_y(\cdot, y_u^\xi) + \xi_y(\cdot, y_u^\xi)\big]p_u^{\xi,\eta} = g_y(\cdot, y_u^\xi, u) + \eta_y(\cdot, y_u^\xi, u).$$

Moreover, there exist positive numbers M and δ such that $|p_u^{\xi,\eta}|_{L^\infty(\Omega)} \leq M$ for all $u \in \mathcal{U}$, $\xi \in \Upsilon_s$ and $\eta \in \Upsilon_c$ with $d_C(\xi,0) + d_C(\xi_y,0) + d_C(\eta_y,0) \leq \delta$.

We denote $\Upsilon := \Upsilon_s \times \Upsilon_c$, and write $\zeta := (\xi, \eta)$ for a generic element of Υ. We endow Υ with the pseudometric $d_\Upsilon : \Upsilon \times \Upsilon \to [0,\infty)$ given by

$$d_\Upsilon(\zeta,\zeta') := d_C(\xi,\xi') + d_C(\xi_y,\xi'_y) + d_C(\eta_y,\eta'_y) + d_C(\eta_u,\eta'_u).$$

5.2 The stability result

We are now ready to state problem (32)-(33) in a precise way. Given $\zeta \in \Upsilon$, problem $\mathcal{P}_\zeta$ is given by

$$\min_{u\in\mathcal{U}} \left\{ \mathcal{J}_\zeta(u) := \int_\Omega \Big[g(x, y_u^\xi, u) + \eta(x, y_u^\xi, u)\Big]\, dx \right\}. \tag{34}$$

Due to the convexity of the cost in the control variable, each problem $\mathcal{P}_\zeta$ has at least one local solution. For each $\zeta \in \Upsilon$, we fix a local minimizer $\hat{u}_\zeta \in \mathcal{U}$ of problem $\mathcal{P}_\zeta$. By the Pontryagin principle, for each $\zeta = (\xi,\eta) \in \Upsilon$, the triple $(\hat{y}_\zeta, \hat{p}_\zeta, \hat{u}_\zeta) := (y_{\hat{u}_\zeta}^\xi, p_{\hat{u}_\zeta}^{\xi,\eta}, \hat{u}_\zeta)$ satisfies the system

$$\left\{\begin{array}{lcl} 0 &=& \mathcal{L}y - f(\cdot,y,u) - \xi(\cdot,y), \\ 0 &=& \mathcal{L}p - H_y(\cdot,y,p,u) + \eta_y(\cdot,y,u) - \xi_y(\cdot,y)p, \\ 0 &\in& H_u(\cdot,y,p) + \eta_u(\cdot,y,u) + N_\mathcal{U}(u). \end{array}\right. \tag{35}$$

As a consequence of Theorem 4.11, we have the following result.

Theorem 5.4. *Let Assumption 2 hold. There exist positive numbers α, α' and c such that*

$$|\hat{y}_\zeta - y_{\bar{u}}|_{L^2(\Omega)} + |\hat{p}_\zeta - p_{\bar{u}}|_{L^2(\Omega)} + |\hat{u}_\zeta - \bar{u}|_{L^1(\Omega)} \leq c d_\Upsilon(\zeta, 0)^{1/k^*}$$

for all $\zeta \in \Upsilon$ such that $|\hat{u}_\zeta - \bar{u}|_{L^1(\Omega)} \leq \alpha$ and $d_\Upsilon(\zeta, 0) \leq \alpha'$.

Proof. By Theorem 4.11, the mapping Φ is strongly Hölder subregular at $(\bar{\psi}, 0)$ with exponent $1/k^*$. Let α_1, α_2 and κ be the positive numbers in the definition of strong subregularity. By Proposition 5.2 and 5.3 there exist positive numbers M and δ_0 such that

$$|y_u^\xi|_{L^\infty(\Omega)} + |p_u^{\xi,\eta}|_{L^\infty(\Omega)} \leq M$$

for all $u \in \mathcal{U}$ and $\zeta \in \Upsilon$ with $d_\Upsilon(\zeta, 0) \leq \delta_0$. Let $K := \bar{\Omega} \times [-M, M]$. By Lemma 5.1, there exists $m \in \mathbb{N}$ such that

$$|\xi(\cdot, y_u^\xi)|_{L^2(\Omega)} \leq \operatorname{meas} \Omega^{\frac{1}{2}} |\xi|_{L^\infty(K)} \leq 2^m \operatorname{meas} \Omega^{\frac{1}{2}} d_C(\xi, 0) \leq 2^m \operatorname{meas} \Omega^{\frac{1}{2}} d_\Upsilon(\zeta, 0)$$

for all $u \in \mathcal{U}$ and $\zeta \in \Upsilon$ with $d_\Upsilon(\zeta, 0) \leq \min\{2^{-m}, \delta_0\}$. Repeating this argument, we can find positive numbers δ and c_0 such that

$$|\xi(\cdot, y_u^\xi)|_{L^2(\Omega)} + |\xi_y(\cdot, y_u^\xi) p_u^{\xi,\eta}|_{L^2(\Omega)} + |\eta_y(\cdot, y_u^\xi, u)|_{L^2 \Omega} + |\eta_u(\cdot, y_u^\xi, u)|_{L^\infty} \leq c_0 d_\Upsilon(\zeta, 0) \tag{36}$$

for all $u \in \mathcal{U}$ and $\zeta \in \Upsilon$ with $d_\Upsilon(\zeta, 0) \leq \delta$. Using Proposition 2.11 and Lemma 5.1, one can find positive numbers α and δ' such that

$$|\hat{y}_\zeta - y_{\bar{u}}|_{L^2(\Omega)} + |\hat{p}_\zeta - p_{\bar{u}}|_{L^2(\Omega)} + |\hat{u}_\zeta - \bar{u}|_{L^1(\Omega)} \leq \alpha_1$$

for all $\zeta \in \Upsilon$ with $|\hat{u}_\zeta - \bar{u}|_{L^1(\Omega)} \leq \alpha$ and $d_\Upsilon(\zeta, 0) \leq \delta'$. Observe that by (35), we have

$$\begin{pmatrix} \xi(\cdot, \hat{y}_\zeta) \\ -\eta_y(\cdot, \hat{y}_\zeta, \hat{u}_\zeta) + \xi_y(\cdot, \hat{y}_\zeta)\hat{p}_\zeta \\ -\eta_u(\cdot, \hat{y}_\zeta, \hat{u}_\zeta) \end{pmatrix} \in \Phi(\hat{y}_\zeta, \hat{p}_\zeta, \hat{u}_\zeta)$$

for all $\zeta \in \Upsilon$. Let $\alpha' := \min\{c_0^{-1}\alpha_2, \delta, \delta'\}$. Then by Hölder subregularity of Φ and (36),

$$|\hat{y}_\zeta - y_{\bar{u}}|_{L^2(\Omega)} + |\hat{p}_\zeta - p_{\bar{u}}|_{L^2(\Omega)} + |u_\zeta - \bar{u}|_{L^1(\Omega)} \leq \kappa c_0^{\frac{1}{k^*}} d_\Upsilon(\zeta, 0)^{\frac{1}{k^*}}$$

for all $\zeta \in \Upsilon$ such that $|\hat{u}_\zeta - \bar{u}|_{L^1(\Omega)} \leq \alpha$ and $d_\Upsilon(\zeta, 0) \leq \alpha'$. The result follows defining $c := \kappa c_0^{\frac{1}{k^*}}$. □

5.3 An application: Tikhonov regularization

In what follows we present an application of the theory derived in the previous chapters, namely the so-called Tikhonov regularization. For a more detailed description and an account of the state of art, the reader is referred to [32, 41, 40]. We derive estimates on the convergence rate of the solution of the regularized problem when the regularization parameter tends to zero. The results that appear in the literature require the so-called structural assumption and positive-definiteness (in some sense) of the second derivative of the objective functional. Using Theorem 4.11, we can obtain this results under weaker assumptions than used in the literature so far. One can compare this results with [32, Theorem 4.4] (where a tracking problem with semilinear elliptic equation is considered) when it comes to stability of the controls. In Section 6, we give more details on how the assumptions in the literature interplay with Assumption 2.

We consider the following family of problems $\{\mathcal{P}_\varepsilon\}_{\varepsilon \geq 0}$.

$$\min_{u \in \mathcal{U}} \left\{ \int_\Omega g(x, y, u)\, dx + \frac{\varepsilon}{2} \int_\Omega u^2\, dx \right\}, \tag{37}$$

subject to

$$\begin{cases} -\operatorname{div}\big(A(x)\nabla y\big) + d(x, y) &= \beta(x) u \quad \text{in} \quad \Omega \\ A(x)\nabla y \cdot \nu + b(x) y &= 0 \quad \text{on} \quad \partial\Omega. \end{cases} \tag{38}$$

For each $\varepsilon > 0$ we fix a local solution $\hat{u}_\varepsilon \in \mathcal{U}$ of problem $\mathcal{P}_\varepsilon$.

Theorem 5.5. *Let Assumption 2 be fulfilled. Then there exist positive constants* α *and* κ *such that*

$$|\hat{u}_\varepsilon - \bar{u}|_{L^1(\Omega)} \le \kappa \varepsilon^{1/k^*} \tag{39}$$

for every $\varepsilon > 0$ *such that* $|\hat{u}_\varepsilon - \bar{u}|_{L^1(\Omega)} \le \alpha$. *If in addition, each* $\hat{u}_\varepsilon$ *is a global solution of problem (37)–(38) then the last claim holds with* $\alpha = +\infty$, *i.e., for every* $\varepsilon > 0$.

Proof. Let $\alpha, \alpha'$and c be the positive numbers in Theorem 5.4. Define $\eta_\varepsilon : \mathbb{R} \to \mathbb{R}$ by $\eta_\varepsilon(u) := \varepsilon u^2/2$ and $\zeta_\varepsilon := (0, \eta_\varepsilon) \in \Upsilon$ for each $\varepsilon > 0$. Note that

$$d_C(\eta_\varepsilon, 0) := \sum_{m=1}^{\infty} \frac{1}{2^m} \frac{\varepsilon m^2/2}{1+\varepsilon m^2/2} = \varepsilon \sum_{m=1}^{\infty} \frac{1}{2^m} \frac{m^2}{2+\varepsilon m^2} \le \varepsilon \sum_{m=1}^{\infty} \frac{m^2}{2^{m+1}} = 3\varepsilon$$

for all $\varepsilon > 0$. Analogously,

$$d_C(\frac{\partial \eta_\varepsilon}{\partial u}, 0) := \sum_{m=1}^{\infty} \frac{1}{2^m} \frac{\varepsilon m}{1+\varepsilon m} \le \varepsilon \sum_{m=1}^{\infty} \frac{m}{2^m} = 2\varepsilon$$

for all $\varepsilon > 0$. We conclude that $d_\Upsilon(\zeta_\varepsilon, 0) \le 5\varepsilon \le \alpha'$ for all $\varepsilon \in (0, \varepsilon_0)$, where $\varepsilon_0 := \alpha'/5$. By Theorem 5.4,

$$|\hat{u}_\varepsilon - \bar{u}|_{L^1(\Omega)} \le 5^{\frac{1}{k^*}} c \varepsilon^{\frac{1}{k^*}}$$

for all $\varepsilon \in (0, \varepsilon_0)$ such that $|\hat{u}_\varepsilon - \bar{u}| \le \alpha$. Let $M > 0$ be a bound for $\mathcal{U}$ in $L^\infty(\Omega)$. We also have we have

$$|\hat{u}_\varepsilon - \bar{u}|_{L^1(\Omega)} \le 2M \le 2M\varepsilon^{-\frac{1}{k^*}} \varepsilon^{\frac{1}{k^*}} \le 2M\varepsilon_0^{-\frac{1}{k^*}} \varepsilon^{\frac{1}{k^*}}$$

for all $\varepsilon \ge \varepsilon_0$. Hence, defining

$$\kappa := \max\left\{5^{\frac{1}{k^*}} c,\, 2M\varepsilon_0^{-1/k^*}\right\},$$

we obtain the first claim.

Let us prove the second claim of the theorem. First we prove that there exists $\varepsilon^* > 0$ such that $|\hat{u}_\varepsilon - \bar{u}|_{L^1(\Omega)} \le \alpha$ for all $\varepsilon \in (0, \varepsilon^*)$. Suppose the opposite. Then there exists a sequence $\{\varepsilon_k\}_{k=1}^\infty$ converging to zero such that $|\hat{u}_{\varepsilon_k} - \bar{u}|_{L^1(\Omega)} > \alpha$ for all $k \in \mathbb{N}$. Since $\mathcal{U}$ is weakly compact in $L^2(\Omega)$, we may assume without loss of generality that $u_{\varepsilon_k} \to u^*$ for some $u^* \in \mathcal{U}$. Since $y_{u_{\varepsilon_k}} \to y_{u^*}$ in $C(\bar{\Omega})$, we obtain that

$$J(u^*) \le \liminf_{k\to\infty} \left[J(u_{\varepsilon_k}) + \frac{\varepsilon_k}{2}|u_{\varepsilon_k}|_{L^2(\Omega)}\right] \le \liminf_{k\to\infty} \left[J(\bar{u}) + \frac{\varepsilon_k}{2}|\bar{u}|_{L^2(\Omega)}\right] = J(\bar{u}).$$

By Remark 4.10, $\bar{u}$ is a strict local solution, therefore $u^* = \bar{u}$. By Proposition 4.1, $u^* = \bar{u}$ is bang-bang. Weak convergence in $L^2(\Omega)$ implies that in $L^1(\Omega)$; consequently, by Lemma 4.2, $u_{\varepsilon_n} \to u^*$ in $L^1(\Omega)$, which is a contradiction. Then the first claim of the theorem implies (39) for all $\varepsilon \in (0, \varepsilon^*)$. For $\varepsilon \ge \varepsilon^*$, (39) remains true if we increase the constant c (if needed) so that $c \ge 2M(\varepsilon^*)^{-1/k^*}$. □

6 Assumptions related to subregularity

In this section, we gather some results concerning Assumption 2, in order to provide sufficient conditions under which it is fulfilled. Furthermore, we analyze related assumptions and their relation between themselves. Recall that $\bar{u} \in \mathcal{U}$ is a local solution of problem (1)–(2). Since $\bar{u} \in \mathcal{U}$ satisfies the variational inequality (9), we have

$$\bar{u}(x) = \begin{cases} b_1(x) & if \quad \sigma_{\bar{u}}(x) > 0 \\[2ex] b_2(x) & if \quad \sigma_{\bar{u}}(x) < 0. \end{cases}$$

We introduce the following extended cone suggested in [5]. For a fixed $\tau > 0$ define

$$C_{\bar{u}}^\tau = \left\{ v \in L^2(\Omega) : v(x) \begin{cases} = 0 & \text{if} \quad |\sigma_{\bar{u}}(x)| > \tau \text{ or } \bar{u}(x) \in (b_1(x), b_2(x)) \\ \ge 0 & \text{if} \quad |\sigma_{\bar{u}}(x)| \le \tau \text{ and } \bar{u}(x) = b_1(x) \\ \le 0 & \text{if} \quad |\sigma_{\bar{u}}(x)| \le \tau \text{ and } \bar{u}(x) = b_2(x) \end{cases} \right\}.$$

We introduce the following modification of Assumption 2.

Assumption 2′. *There exist positive numbers α_0 and γ_0 such that*

$$\int_\Omega \sigma_{\bar u}(u-\bar u)\,dx + \Lambda(u-\bar u) \geq \gamma_0 |u-\bar u|_{L^1(\Omega)}^{k^*+1},$$

for all $u \in \mathcal{U}$ with $u - \bar u \in C_{\bar u}^\tau \cap \mathbb{B}_{L^1(\Omega)}(\bar u; \alpha_0)$.

This assumption is seemingly weaker than Assumption 2. However, we will prove that the two assumptions are equivalent. Before that, for technical purposes, we introduce the bilinear form $\Gamma : L^2(\Omega) \times L^2(\Omega) \to \mathbb{R}$ given by

$$\Gamma(v_1, v_2) := \frac{1}{2}\int_\Omega \Big[\pi_{v_1} v_2 + \pi_{v_2} v_1\Big]\,dx. \tag{40}$$

The bilinear form is particularly useful because of the following property.

$$\Lambda(v_1+v_2) = \Gamma(v_1,v_1) + 2\Gamma(v_1,v_2) + \Gamma(v_2,v_2) \quad \forall v_1, v_2 \in L^2(\Omega). \tag{41}$$

We will require the following technical lemma.

Lemma 6.1. *For every positive number M, there exits a positive number c such that*

$$|\Gamma(v_1,v_2)| \leq c|v_1|_{L^1(\Omega)}^{1/2}|v_2|_{L^1(\Omega)}$$

for all $v_1, v_2 \in \mathbb{B}_{L^\infty}(0;M)$.

Proof. By Proposition 3.6, there exist $c_1, c_2 > 0$ such that $|\pi_v|_{L^\infty(\Omega)} \leq c_1|v|_{L^2(\Omega)}$ and $|\pi_v|_{L^2(\Omega)} \leq c_2|v|_{L^1(\Omega)}$ for all $v \in L^2(\Omega)$. Let $M > 0$ be arbitrary. Observe that

$$\Big|\int_\Omega \pi_{v_1} v_2\,dx\Big| \leq |\pi_{v_1}|_{L^\infty(\Omega)}|v_2|_{L^1(\Omega)} \leq c_1 M^{\frac12}|v_1|_{L^1(\Omega)}^{\frac12}|v_2|_{L^1(\Omega)},$$

and that

$$\Big|\int_\Omega \pi_{v_2} v_1\,dx\Big| \leq |\pi_{v_2}|_{L^2(\Omega)}|v_1|_{L^2(\Omega)} \leq c_2 M^{\frac12}|v_1|_{L^1(\Omega)}^{\frac12}|v_2|_{L^1(\Omega)}$$

for all $v_1, v_2 \in \mathbb{B}_{L^\infty}(0;M)$. There result follows defining $c := 2^{-1}(c_1+c_2)M^{\frac12}$. □

Proposition 6.2. *Assumptions 2 and 2′ are equivalent.*

Proof. Let α_0 and γ_0 be the numbers in Assumption 2′. Let $u \in \mathcal{U}$ and define

$$v_1(x) := \begin{cases} u(x) - \bar u(x) & if \quad |\sigma_{\bar u}(x)| \leq \tau \\ 0 & if \quad |\sigma_{\bar u}(x)| > \tau, \end{cases}$$

and

$$v_2(x) := \begin{cases} 0 & if \quad |\sigma_{\bar u}(x)| \leq \tau \\ u(x) - \bar u(x) & if \quad |\sigma_{\bar u}(x)| > \tau. \end{cases}$$

Clearly $v_1 \in C_{\bar u}^\tau$ and $v_1 + v_2 = u - \bar u$. Let M be a bound for $\mathcal{U}$ in $L^\infty(\Omega)$, and let c be the positive number in Lemma 6.1 corresponding to $2M$. By Assumption 2′,

$$\begin{aligned}\int_\Omega \sigma_{\bar u}(u-\bar u)\,dx &= \int_\Omega \sigma_{\bar u} v_1\,dx + \int_{|\sigma_{\bar u}|>\tau} \sigma_{\bar u} v_2\,dx \\ &= \int_\Omega \sigma_{\bar u} v_1\,dx + \Lambda(v_1) - \Lambda(v_1) + \int_{|\sigma_{\bar u}|>\tau} \sigma_{\bar u} v_2\,dx \\ &\geq \gamma_0|v_1|^{k+1} + \tau|v_2|_{L^1(\Omega)} - \Lambda(v_1),\end{aligned}$$

and

$$\begin{aligned}\Lambda(u-\bar u) &= \Lambda(v_1) + 2\Gamma(v_1,v_2) + \Lambda(v_2)\\ &\geq \Lambda(v_1) - 2c|v_1|_{L^1(\Omega)}^{1/2}|v_2|_{L^1(\Omega)} - c|v_2|_{L^1(\Omega)}^{1/2}|v_2|_{L^1(\Omega)}\\ &\geq \Lambda(v_1) - 3c|v_2|_{L^1(\Omega)}|u-\bar u|_{L^1(\Omega)}^{1/2}\end{aligned}$$

for $u \in \mathcal{U}$ with $|u-\bar u|_{L^1(\Omega)} \leq \alpha_0$. Thus

$$\begin{aligned}\int_\Omega \sigma_{\bar u}(u-\bar u)\,dx + \Lambda(u-\bar u) &\geq \gamma_0|v_1|^{k+1} + \tau|v_2|_{L^1(\Omega)} - 3c|v_2|_{L^1(\Omega)}|u-\bar u|_{L^1(\Omega)}^{1/2}\\ &= \gamma_0|v_1|^{k+1} + |v_2|_{L^1(\Omega)}\Big(\tau - 3c|u-\bar u|_{L^1(\Omega)}^{1/2}\Big)\end{aligned}$$

for $u \in \mathcal{U}$ with $|u-\bar u|_{L^1(\Omega)} \leq \alpha_0$. Now, by the reverse triangle inequality and Bernoulli's inequality (consider without loss of generality $u \neq \bar u$)

$$\begin{aligned}|v_1|_{L^1(\Omega)}^{k+1} &= |(u-\bar u) - v_2|_{L^1(\Omega)}^{k+1} \geq \Big(|u-\bar u|_{L^1(\Omega)} - |v_2|_{L^1(\Omega)}\Big)^{k+1}\\ &= |u-\bar u|_{L^1(\Omega)}^{k+1}\Big(1 - \frac{|v_2|_{L^1(\Omega)}}{|u-\bar u|_{L^1(\Omega)}}\Big)^{k+1} \geq |u-\bar u|_{L^1(\Omega)}^{k+1}\Big(1-(k+1)\frac{|v_2|_{L^1(\Omega)}}{|u-\bar u|_{L^1(\Omega)}}\Big)\\ &= |u-\bar u|_{L^1(\Omega)}^{k+1} - (k+1)|u-\bar u|_{L^1(\Omega)}^{k}|v_2|_{L^1(\Omega)}.\end{aligned}$$

Consequently,

$$\begin{aligned}\int_\Omega \sigma_{\bar u}(u-\bar u)\,dx + \Lambda(u-\bar u) &\geq \gamma_0|v_1|^{k+1} + |v_2|_{L^1(\Omega)}\Big(\tau - 3c|u-\bar u|_{L^1(\Omega)}^{1/2}\Big)\\ &\geq \gamma_0|u-\bar u|_{L^1(\Omega)}^{k+1} - \gamma_0(k+1)|u-\bar u|_{L^1(\Omega)}^{k}|v_2|_{L^1(\Omega)} + |v_2|_{L^1(\Omega)}\Big(\tau - 3c|u-\bar u|_{L^1(\Omega)}^{1/2}\Big)\\ &\geq \gamma_0|u-\bar u|_{L^1(\Omega)}^{k+1} + |v_2|_{L^1(\Omega)}\Big(\tau - \gamma_0(k+1)|u-\bar u|_{L^1(\Omega)}^{k} - 3c|u-\bar u|_{L^1(\Omega)}^{1/2}\Big).\end{aligned}$$

Choosing α small enough, one can ensure

$$\begin{aligned}\int_\Omega \sigma_{\bar u}(u-\bar u)\,dx + \Lambda(u-\bar u) &\geq \gamma_0|u-\bar u|_{L^1(\Omega)}^{k+1} + |v_2|_{L^1(\Omega)}\Big(\tau - \gamma_0(k+1)|u-\bar u|_{L^1(\Omega)}^{k} - 3c|u-\bar u|_{L^1(\Omega)}^{1/2}\Big)\\ &\geq \gamma_0|u-\bar u|_{L^1(\Omega)}^{k+1} + \frac{\tau}{2}|v_2|_{L^1(\Omega)} \geq \gamma_0|u-\bar u|_{L^1(\Omega)}^{k+1}\end{aligned}$$

for all $u \in \mathcal{U}$ with $|u-\bar u|_{L^1(\Omega)} \leq \alpha$.

□

Proposition 6.2 allows to split Assumption 2 in two parts, as it follows in the next theorem.

Theorem 6.3. *Let there exist numbers $\mu_1, \mu_2 \in \mathbb{R}$ and $\alpha > 0$ such that*

$$\int_\Omega \sigma_{\bar u} v\,dx \geq \mu_1 |v|_{L^1(\Omega)}^{k^*+1} \tag{42}$$

and

$$\Lambda(v) \geq \mu_2 |v|_{L^1(\Omega)}^{k^*+1} \tag{43}$$

for every $v \in (\mathcal{U}-\bar u) \cap C_{\bar u}^{\tau} \cap \mathbb{B}_{L^1(\Omega)}(\bar u;\alpha)$. If $\mu_1+\mu_2>0$, then Assumption 2 is fulfilled, hence the optimality mapping Φ (see (22)) of problem (1)–(2) is strongly Hölder subregular with exponent $\lambda = 1/k^*$ at the reference point $(\bar y, \bar p, \bar u)$.

The proof consists of summation of (42) and (43) and utilization of Proposition 6.2 and Theorem 4.11.

The splitting of Assumption 2 has the advantage that the inequalities in (42) and (43) can be analyzed separately. The next proposition is related to (42).

The following assumption has become standard in the literature on PDE optimal control problems with bang-bang controls, see, e.g., [9, 12, 34, 42].

Assumption 3. *There exists a positive number* μ_0 *such that*

$$meas\,\{x \in \Omega : |\sigma_{\bar{u}}(x)| \leq \varepsilon\} \leq \mu_0 \varepsilon^{\frac{1}{k^*}} \quad \forall \varepsilon > 0.$$

Proposition 6.4. *The following statements hold.*

(i) If Assumption 3 is fulfilled then there exists $\mu_1 > 0$ *such that (42) holds for every* $v \in \mathcal{U} - \bar{u}$.

(ii) Suppose there exists $\nu > 0$ *such that* $b_2(x) - b_1(x) \geq \nu$ *for a.e.* $x \in \Omega$. *If (42) holds for every* $v \in \mathcal{U} - \bar{u}$ *then Assumption 3 is fulfilled.*

Proof. The proof of the first claim follows [34, Proposition 3.1], see also [9, Proposition 2.7]. It has been also proved several times in the literature on ordinary differential equations in a somewhat stronger form; see, e.g., [1, 28, 33, 37].

Let us prove the second claim. For each $\varepsilon > 0$, define

$$u_\varepsilon(x) := \begin{cases} \bar{u}(x) & \text{if} \qquad\qquad\qquad\qquad |\sigma_{\bar{u}}(x)| > \varepsilon \\ b_1(x) & \text{if} \quad |\sigma_{\bar{u}}(x)| \leq \varepsilon \quad \text{and} \quad \bar{u}(x) \in \Big[\dfrac{b_1(x)+b_2(x)}{2}, b_2(x)\Big] \\ b_2(x) & \text{if} \quad |\sigma_{\bar{u}}(x)| \leq \varepsilon \quad \text{and} \quad \bar{u}(x) \in \Big[b_1(x), \dfrac{b_1(x)+b_2(x)}{2}\Big). \end{cases}$$

Clearly each u_ε belongs to $\mathcal{U}$, and

$$|u_\varepsilon(x) - \bar{u}(x)| \geq \frac{1}{2}|b_2(x) - b_1(x)| \tag{44}$$

for a.e $x \in \{s \in \Omega : |\sigma_{\bar{u}}(s)| \leq \epsilon\}$. From (42) we have

$$\mu_1 \Big(\int_{|\sigma_{\bar{u}}| \leq \varepsilon} |u_\varepsilon - \bar{u}|\, dx \Big)^{k+1} \leq \int_{|\sigma_{\bar{u}}| \leq \varepsilon} \sigma_{\bar{u}} (u_\varepsilon - \bar{u})\, dx \leq \varepsilon \int_{|\sigma_{\bar{u}}| \leq \varepsilon} |u_\varepsilon - \bar{u}|\, dx.$$

This implies

$$\int_{|\sigma_{\bar{u}}| \leq \varepsilon} |u_\varepsilon - \bar{u}|\, dx \leq \mu_1^{-\frac{1}{k}} \varepsilon^{\frac{1}{k}}. \tag{45}$$

Using (44) and (45) we obtain that

$$\begin{aligned} \text{meas}\,\{x \in \Omega : |\sigma_{\bar{u}}(x)| \,\leq\, \varepsilon\} &= \frac{1}{\nu} \int_{|\sigma_{\bar{u}}| \leq \varepsilon} \nu\, dx \,\leq\, \frac{1}{\nu} \int_{|\sigma_{\bar{u}}| \leq \varepsilon} |b_2 - b_1|\, dx \,\leq\, \frac{2}{\nu} \int_{|\sigma_{\bar{u}}| \leq \varepsilon} |u_\varepsilon - \bar{u}|\, dx \\ &\leq 2(\mu_1)^{-\frac{1}{k}} \nu^{-1} \varepsilon^{\frac{1}{k}}. \end{aligned}$$

Thus Assumption 3 is fulfilled with $\mu_0 := 2(\mu_1)^{-\frac{1}{k}} \nu^{-1}$. □

References

[1] W. Alt, C. Schneider, and M. Seydenschwanz. Regularization and implicit Euler discretization of linear-quadratic optimal control problems with bang-bang solutions. *Appl. Math. Comput.*, 287/288:104–124, 2016.

[2] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems.* Springer Series in Operations Research. Springer-Verlag, New York, 2000.

[3] R. S. Burachik and V. Jeyakumar. A simple closure condition for the normal cone intersection formula. *Proc. Amer. Math. Soc.*, 133(6):1741–1748, 2005.

[4] E. Casas. Boundary control of semilinear elliptic equations with pointwise state constraints. *SIAM J. Control Optim.*, 31(4):993–1006, 1993.

[5] E. Casas. Second order analysis for bang-bang control problems of PDEs. *SIAM J. Control Optim.*, 50(4):2355–2372, 2012.

[6] E. Casas, J. C. de los Reyes, and F. Tröltzsch. Sufficient second-order optimality conditions for semilinear control problems with pointwise state constraints. *SIAM J. Optim.*, 19(2):616–643, 2008.

[7] E. Casas, M. Mateos, and A. Rösch. Analysis of control problems of nonmontone semilinear elliptic equations. *ESAIM Control Optim. Calc. Var.*, 26:Paper No. 80, 21, 2020.

[8] E. Casas and F. Tröltzsch. On optimal control problems with controls appearing nonlinearly in an elliptic state equation. *SIAM J. Control Optim.*, 58(4):1961–1983, 2020.

[9] E. Casas, D. Wachsmuth, and G. Wachsmuth. Sufficient second-order conditions for bang-bang control problems. *SIAM J. Control Optim.*, 55(5):3066–3090, 2017.

[10] E. Casas, D. Wachsmuth, and G. Wachsmuth. Second-order analysis and numerical approximation for bang-bang bilinear control problems. *SIAM J. Control Optim.*, 56(6):4203–4227, 2018.

[11] R. Cibulka, A. L. Dontchev, and A. Y. Kruger. Strong metric subregularity of mappings in variational analysis and optimization. *J. Math. Anal. Appl.*, 457(2):1247–1282, 2018.

[12] K. Deckelnick and M. Hinze. A note on the approximation of elliptic control problems with bang-bang controls. *Comput. Optim. Appl.*, 51(2):931–939, 2012.

[13] A. Domínguez Corella, M. Quincampoix, and V. M. Veliov. Strong bi-metric regularity in affine optimal control problems. *Pure Appl. Funct. Anal.*, 6(6):1119–1137, 2021.

[14] A. Domínguez Corella and V. M. Veliov. Hölder regularity in bang-bang type affine optimal control problems. In *Large-scale scientific computing*, volume 13127 of *Lecture Notes in Comput. Sci.*, pages 306–313. Springer, Cham, 2021.

[15] A. L. Dontchev, I. V. Kolmanovsky, M. I. Krastanov, V. M. Veliov, and P. T. Vuong. Approximating optimal finite horizon feedback by model predictive control. *Systems Control Lett.*, 139:104666, 9, 2020.

[16] A. L. Dontchev and R. T. Rockafellar. *Implicit functions and solution mappings*. Springer Monographs in Mathematics. Springer, Dordrecht, 2009. A view from variational analysis.

[17] W. H. Fleming and R. W. Rishel. *Deterministic and Stochastic Optimal Control*. Springer-Verlag, Berlin-New York, 1975.

[18] R. Griesse. Lipschitz stability of solutions to some state-constrained elliptic optimal control problems. *Z. Anal. Anwend.*, 25(4):435–455, 2006.

[19] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1985.

[20] R. Haller-Dintelmann, C. Meyer, J. Rehberg, and A. Schiela. Hölder continuity and optimal control for nonsmooth elliptic problems. *Appl. Math. Optim.*, 60(3):397–428, 2009.

[21] M. Hinze and C. Meyer. Stability of semilinear elliptic optimal control problems with pointwise state constraints. *Comput. Optim. Appl.*, 52(1):87–114, 2012.

[22] B. T. Kien, N. Q. Tuan, C.-F. Wen, and J.-C. Yao. L^∞-stability of a parametric optimal control problem governed by semilinear elliptic equations. *Appl. Math. Optim.*, 84(1):849–876, 2021.

[23] D. Luenberger. *Optimization by Vector Space Methods*. Wiley-Interscience, 1969.

[24] K. Malanowski and F. Tröltzsch. Lipschitz stability of solutions to parametric optimal control for elliptic equations. *Control Cybernet.*, 29(1):237–256, 2000.

[25] B. S. Mordukhovich and T. T. A. Nghia. Full Lipschitzian and Hölderian stability in optimization with applications to mathematical programming and optimal control. *SIAM J. Optim.*, 24(3):1344–1381, 2014.

[26] J. R. Munkres. *Topology: a first course*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1975.

[27] R. Nittka. Regularity of solutions of linear second order elliptic and parabolic boundary value problems on Lipschitz domains. *J. Differential Equations*, 251(4-5):860–880, 2011.

[28] N. P. Osmolovskii and V. M. Veliov. Metric sub-regularity in optimal control of affine problems with free end state. *ESAIM Control Optim. Calc. Var.*, 26:Paper No. 47, 19, 2020.

[29] N. P. Osmolovskii and V. M. Veliov. On the regularity of Mayer-type affine optimal control problems. In *Large-scale scientific computing*, volume 11958 of *Lecture Notes in Comput. Sci.*, pages 56–63. Springer, Cham, [2020] ©2020.

[30] F. Pörner and D. Wachsmuth. An iterative Bregman regularization method for optimal control problems with inequality constraints. *Optimization*, 65(12):2195–2215, 2016.

[31] F. Pörner and D. Wachsmuth. Tikhonov regularization of optimal control problems governed by semi-linear partial differential equations. *Math. Control Relat. Fields*, 8(1):315–335, 2018.

[32] F. Pörner and D. Wachsmuth. Tikhonov regularization of optimal control problems governed by semi-linear partial differential equations. *Math. Control Relat. Fields*, 8(1):315–335, 2018.

[33] J. Preininger, T. Scarinci, and V. M. Veliov. On the regularity of linear-quadratic optimal control problems with bang-bang solutions. In *Large-scale scientific computing*, volume 10665 of *Lecture Notes in Comput. Sci.*, pages 237–245. Springer, Cham, 2018.

[34] N. T. Qui and D. Wachsmuth. Stability for bang-bang control problems of partial differential equations. *Optimization*, 67(12):2157–2177, 2018.

[35] N. T. Qui and D. Wachsmuth. Full stability for a class of control problems of semilinear elliptic partial differential equations. *SIAM J. Control Optim.*, 57(4):3021–3045, 2019.

[36] S. M. Robinson. Generalized equations and their solutions. I. Basic theory. *Math. Programming Stud.*, (10):128–141, 1979. Point-to-set maps and mathematical programming.

[37] M. Seydenschwanz. Convergence results for the discrete regularization of linear-quadratic control problems with bang-bang solutions. *Comput. Optim. Appl.*, 61(3):731–760, 2015.

[38] G. Stampacchia. Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. *Ann. Inst. Fourier (Grenoble)*, 15(fasc. 1):189–258, 1965.

[39] F. Tröltzsch. *Optimal control of partial differential equations*, volume 112 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.

[40] N. von Daniels. Tikhonov regularization of control-constrained optimal control problems. *Comput. Optim. Appl.*, 70(1):295–320, 2018.

[41] D. Wachsmuth and G. Wachsmuth. Regularization error estimates and discrepancy principle for optimal control problems with inequality constraints. *Control Cybernet.*, 40(4):1125–1158, 2011.

[42] G. Wachsmuth and D. Wachsmuth. Convergence and regularization results for optimal control problems with sparsity functional. *ESAIM Control Optim. Calc. Var.*, 17(3):858–886, 2011.

Chapter 5

New Assumptions for Stability Analysis in Elliptic Optimal Control Problems

This chapter consists of the paper:

- Casas, Eduardo and Domínguez Corella, Alberto and Jork, Nicolai. New assumptions for stability analysis in elliptic optimal control problems.

It is conditionally accepted (after a minor revision) in SIAM Journal on Control and Optimization. The author of this thesis proposed the idea for the paper, and gave an outline of the main result's proof; he wrote a first version of the paper (including a sketch of sections 1-4) that was later corrected and enriched by the other authors.

NEW ASSUMPTIONS FOR STABILITY ANALYSIS IN ELLIPTIC OPTIMAL CONTROL PROBLEMS*

EDUARDO CASAS†, ALBERTO DOMÍNGUEZ CORELLA‡, AND NICOLAI JORK§

Abstract. This paper is dedicated to the stability analysis of the optimal solutions of a control problem associated with a semilinear elliptic equation. The linear differential operator of the equation is neither monotone nor coercive due to the presence of a convection term. The control appears only linearly, or even it can not appear in a explicit form in the objective functional. Under new assumptions, we prove Lipschitz stability of the optimal controls and associated states with respect to perturbations in the equation and the objective functional as well as with respect to the Tikhonov regularization parameter.

AMS subject classifications. 35J61, 49J20, 49K20, 49K40

Key words. Semilinear elliptic equations, optimality conditions, stability analysis, Tikhonov regularization

1. Introduction.

In this paper, we study the following optimal control problem

$$\text{(P)} \quad \min_{u\in\mathcal{U}_{ad}} J(u) := \int_\Omega L(x, y_u(x), u(x))\,\mathrm{d}x,$$

where $\mathcal{U}_{ad} = \{u \in L^2(\Omega) : u_a \le u(x) \le u_b \text{ for a.a. } x \in \Omega\}$, $-\infty < u_a < u_b < +\infty$. Here, y_u denotes the solution of the semilinear elliptic equation:

$$\text{(1.1)} \qquad \left\{\begin{array}{l} -\operatorname{div}\big(A(x)\nabla y\big) + b(x)\cdot\nabla y + f(x,y) = u \text{ in } \Omega,\\ y = 0 \text{ on } \Gamma. \end{array}\right.$$

Assumptions on the data of the control problem (P) will be given below. The aim of this paper is to prove stability results for the local minimizers of (P) with respect to perturbations in the data of the control problem. There are quite a few previous papers devoted to this issue [14], [15], [16], [17], just to mention some of them. In all these cases, the second derivative of L with respect to u is bounded from below by a positive constant. This is the case where the Tikhonov term is involved in the objective functional. Under this condition and assuming sufficient second order optimality conditions (SSOC), the Lipschitz stability of the optimal controls is proved. Here, we assume that u appears linearly in $L(x, y, u)$ or even it does not appear at all. Therefore, the previous results do not apply. In this case, under (SSOC) for optimality, Lipschitz stability of the optimal states can be proved; see [7]. In Section 4, we obtain analogous estimates for the optimal states replacing (SSOC) by a weaker condition; see (3.13). It is weaker because (SSOC) implies our assumption, but they are not equivalent. In addition, our assumption implies strict local optimality of the control; see Theorem 3.5.

*The first author was supported by MCIN/ AEI/10.13039/501100011033/ under research project PID2020-114837GB-I00. The second and third authors were supported by the Austrian Science Foundation (FWF) under grant No I4571.

†Departamento de Matemática Aplicada y Ciencias de la Computación, E.T.S.I. Industriales y de Telecomunicación, Universidad de Cantabria, Santander, España, eduardo.casas@unican.es

‡Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, alberto.corella@tuwien.ac.at

§Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, nicolai.jork@tuwien.ac.at

In order to prove stability of the optimal controls when they are not explicitly involved in the objective functional, besides (SSOC) an additional structural hypothesis is usually assumed. This situation was studied in [21], where the authors proved Lipschitz stability of the control with respect to linear perturbations simultaneously appearing in the state equation and the objective functional. The drawback is that the additional hypothesis is satisfied only by bang-bang controls. Here, we obtain analogous estimates changing the mentioned assumption by a weaker one, see (5.2). Though this second assumption (5.2) is stronger than (3.13), it can be satisfied by optimal controls independently if they are bang-bang or not. Moreover our assumption (5.2) is satisfied if the (SSOC) and the additional hypothesis are assumed.

Finally, under the assumption (5.2), Lipschitz stability is established for the optimal states with respect to simultaneous perturbations in the equations and in the objective functional with respect to the state and the control, and with respect to the Tikhonov regularization parameter. The stability with respect the Tikhonov regularization has been studied in [7] and [20]. In [7], Hölder stability of the states is proved. In [20], stability of the control is proved under (SSOC) and the structural assumption. The reader is also referred to [23], [24], [25] for the case of linear partial differential equations.

In this paper, besides providing some new sufficient conditions for Lipschitz stability for the optimal control and associated states, we deal with a semilinear elliptic state equation that is neither monotone nor coercive. Though some crucial results for this state equation are taken from [6], some estimates have been proved that are not available in the literature.

The plan of this paper is as follows. In Section 2, we analyze the state equation. First, we establish some properties of the linear differential operator of the state equation, and the full semilinear equation is analyzed in the second part of the section. The control problem (P) is studied in Section 3. We prove that our assumption (3.13) is a sufficient condition for strong local optimality. Session 4 is dedicated to the proof of Lipschitz stability of the optimal states. In Section 5 we introduce the stronger condition (5.2) replacing (3.13) that allows us to establish the Lipschitz stability of the optimal controls. Finally, in Section 6, the Tikhonov regularization is considered.

2. Analysis of the partial differential equation. In this section we analyze the equation (1.1). We split the section in two parts. In the first part, we establish the results concerning the linear operator of the elliptic equation. In the second subsection, the nonlinear equation will be studied.

2.1. Analysis of the linear differential operator. We define the differential operator $\mathcal{A} : H_0^1(\Omega) \longrightarrow H^{-1}(\Omega)$ by

$$\mathcal{A}y = -\text{div}\big(A(x)\nabla y\big) + b(x) \cdot \nabla y.$$

The following assumptions are supposed to hold throughout the paper. They ensure that the mathematical objects under consideration are well defined.

ASSUMPTION 2.1. *The following statements are fulfilled.*

(i) The set $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, is a bounded domain with a Lipschitz boundary Γ. The mapping $A : \Omega \longrightarrow \mathbb{R}^{n\times n}$ is measurable and bounded in Ω, and there exists $\Lambda_A > 0$ such that $\xi^\top A(x)\xi \geq \Lambda_A|\xi|^2$ for a.e. $x \in \Omega$ and all $\xi \in \mathbb{R}^n$.

(ii) We assume that $b \in L^p(\mathbb{R}^n)$ with $p \geq 3$ if $n = 3$ and $p > 2$ arbitrary if $n = 2$.

Under theses assumptions it is known that $\mathcal{A} : H_0^1(\Omega) \to H^{-1}(\Omega)$ is an isomorphism despite the fact that the operator is neither coercive nor monotone; see [6], [13, Theorem 8.3], [22]. The following identity is satisfied

$$\langle \mathcal{A}y, z\rangle = \int_\Omega A\nabla y \cdot \nabla z \,\mathrm{d}x + \int_\Omega b \cdot \nabla y z \,\mathrm{d}x \quad \forall y, z \in H_0^1(\Omega),$$

where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between $H^{-1}(\Omega)$ and $H_0^1(\Omega)$.

Along this paper we will set

$$\|y\|_{H_0^1(\Omega)} = \left(\int_\Omega |\nabla y(x)|^2 \,\mathrm{d}x\right)^{\frac{1}{2}}.$$

The next lemma states some properties of $\mathcal{A}$ that will be used later.

Lemma 2.2. *The following statements are fulfilled:*

(i) There exists a constant $C_{\Lambda_A,b}$ such that Gårding's inequality holds

$$\langle \mathcal{A}y, y\rangle \geq \frac{\Lambda_A}{4}\|y\|^2_{H_0^1(\Omega)} - C_{\Lambda_A,b}\|y\|^2_{L^2(\Omega)} \quad \forall y \in H_0^1(\Omega). \tag{2.1}$$

(ii) Let $a \in L^\infty(\Omega)$ be a nonnegative function and $h \in H^{-1}(\Omega)$. If $y \in H_0^1(\Omega)$ satisfies $\mathcal{A}y + ay = h$ and h is a nonnegative linear form, then y is a nonnegative function as well.

(iii) Let a be as above and $h \in L^r(\Omega)$ with $r > \frac{n}{2}$. Then, the solution y of the above equation belongs to $H_0^1(\Omega) \cap C(\bar\Omega)$. Moreover, there exists a constant C_r independent of a and h such that

$$\|y\|_{H_0^1(\Omega)} + \|y\|_{C(\bar\Omega)} \leq C_r\|h\|_{L^r(\Omega)}. \tag{2.2}$$

Proof. The proof of (2.1) can be found in [6]; see also [13, Lemma 8.4]. For the proof of (ii) the reader is referred again to [6] and [13, Theorem 8.1]. The $H_0^1(\Omega)\cap C(\bar\Omega)$ regularity of y for functions $h \in L^r(\Omega)$ is well known; see [13, Lemma 8.31]. It remains to prove the estimates (2.2) for a constant C_r independent of h and a. Let us denote by $y_{a,h} \in H_0^1(\Omega)\cap C(\bar\Omega)$ the solution of $\mathcal{A}y + ay = h$. With $y_{0,h}$ we denote the solution corresponding to $a \equiv 0$. Then, the estimate $\|y_{0,h}\|_{C(\bar\Omega)} \leq C\|h\|_{L^r(\Omega)}$ is well known for a constant C depending on r, but independent of h. Let us write $h = h^+ - h^-$. From (ii) we know that $y_{a,h^+} \geq 0$ and $y_{a,h^-} \geq 0$. Now, since $\mathcal{A}(y_{a,h^+} - y_{0,h^+}) + a(y_{a,h^+} - y_{0,h^+}) = -ay_{0,h^+}$, again by item (ii), we obtain $0 \leq y_{a,h^+} \leq y_{0,h^+}$ and consequently $\|y_{a,h^+}\|_{C(\bar\Omega)} \leq \|y_{0,h^+}\|_{C(\bar\Omega)}$. Analogously, by the same argument $0 \leq y_{a,h^-} \leq y_{0,h^-}$ and consequently $\|y_{a,h^-}\|_{C(\bar\Omega)} \leq \|y_{0,h^-}\|_{C(\bar\Omega)}$. Therefore,

$$\begin{aligned}\|y_{a,h}\|_{C(\bar\Omega)} &\leq \|y_{a,h^+}\|_{C(\bar\Omega)} + \|y_{a,h^-}\|_{C(\bar\Omega)} \leq \|y_{0,h^+}\|_{C(\bar\Omega)} + \|y_{0,h^-}\|_{C(\bar\Omega)}\\ &\leq C\Big(\|h^+\|_{L^r(\Omega)} + \|h^-\|_{L^r(\Omega)}\Big) \leq 2C\|h\|_{L^r(\Omega)},\end{aligned}$$

where C is independent of a and h. To prove the corresponding estimate in $H_0^1(\Omega)$ we use Gårding's inequality (2.1) and the above estimate:

$$\begin{aligned}&\frac{\Lambda_A}{4}\|y_{a,h}\|^2_{H_0^1(\Omega)} \leq \langle \mathcal{A}y_{a,h}, y_{a,h}\rangle + C_{\Lambda_A,b}\|y_{a,h}\|^2_{L^2(\Omega)}\\ &\leq \langle \mathcal{A}y_{a,h}, y_{a,h}\rangle + \int_\Omega a y_{a,h}^2 \,\mathrm{d}x + C_{\Lambda_A,b}\|y_{a,h}\|^2_{L^2(\Omega)}\\ &= \int_\Omega h y_{a,h} \,\mathrm{d}x + C_{\Lambda_A,b}\|y_{a,h}\|^2_{L^2(\Omega)} \leq |\Omega|^{\frac{r-1}{r}}\|h\|_{L^r(\Omega)}\|y_{a,h}\|_{C(\bar\Omega)} + C_{\Lambda_A,b}|\Omega|\|y_{a,h}\|^2_{C(\bar\Omega)}\\ &\leq 2C\Big(|\Omega|^{\frac{r-1}{r}} + 2CC_{\Lambda_A,b}|\Omega|\Big)\|h\|^2_{L^r(\Omega)},\end{aligned}$$

∎

where $|\Omega|$ denotes the Lebesgue measure of Ω. Since the above constants are independent of a and h, the inequality completes the proof of (2.2).

Now, we consider the adjoint operator $\mathcal{A}^* : H_0^1(\Omega) \to H^{-1}(\Omega)$ of $\mathcal{A}$. Since $\mathcal{A}$ is an isomorphism, $\mathcal{A}^*$ is also an isomorphism as well. It is obvious that $\mathcal{A}^*\varphi = -\mathrm{div}\big(A^\top \nabla\varphi\big) - \mathrm{div}\big(\varphi b\big)$. The operator $\mathcal{A}^*$ satisfies the same properties established in Lemma 2.2. Indeed, the Gårding's inequality is consequence of (2.1) and the identity $\langle \mathcal{A}^*\varphi, \varphi\rangle = \langle \mathcal{A}\varphi, \varphi\rangle$. The proof of the estimate (2.2) is the same for the operator $\mathcal{A}^*$. We only prove the statement (ii). Let $h \in H^{-1}(\Omega)$ be a nonnegative linear form. This means that $\langle h, y\rangle \geq 0$ for every nonnegative function $y \in H_0^1(\Omega)$. Let $\varphi \in H_0^1(\Omega)$ satisfy $\mathcal{A}^*\varphi + a\varphi = h$. Now, given a nonnegative function $w \in L^2(\Omega)$ we take $y \in H_0^1(\Omega)$ satisfying $\mathcal{A}y + ay = w$. By Lemma 2.2-(ii) we have that $y \geq 0$. Then, we obtain

$$\int_\Omega w\varphi\,dx = \langle \mathcal{A}y + ay, \varphi\rangle = \langle \mathcal{A}^*\varphi + a\varphi, y\rangle = \langle h, y\rangle \geq 0.$$

Since w is an arbitrary nonnegative function of $L^2(\Omega)$, this inequality yields $\varphi \geq 0$.

We finish this subsection by proving an $L^s(\Omega)$ estimate.

LEMMA 2.3. *Assume that $s \in [1, \frac{n}{n-2})$, s' is its conjugate, and let $a \in L^\infty(\Omega)$ be a nonnegative function. Then, there exists a constant $C_{s'}$ independent of a such that*

$$\left\{\begin{array}{l} \|y_h\|_{L^s(\Omega)} \leq C_{s'}\|h\|_{L^1(\Omega)}, \\ \|\varphi_h\|_{L^s(\Omega)} \leq C_{s'}\|h\|_{L^1(\Omega)}, \end{array}\right. \quad \forall h \in H^{-1}(\Omega)\cap L^1(\Omega), \tag{2.3}$$

where y_h and φ_h satisfy the equations $\mathcal{A}y_h + ay_h = h$ and $\mathcal{A}^*\varphi_h + a\varphi_h = h$, respectively, and $C_{s'}$ is given by (2.2) with $r = s'$.

Proof. We prove the estimate (2.3) for φ_h, the proof being identical for y_h. First we observe that $H_0^1(\Omega) \subset L^{\frac{2n}{n-2}}(\Omega) \subset L^{\frac{n}{n-2}}(\Omega)$, hence $\varphi_h \in L^s(\Omega)$. As a consequence we obtain that $|\varphi_h|^{s-1}\mathrm{sign}(\varphi_h) \in L^{s'}(\Omega)$. Moreover, $s < \frac{n}{n-2}$ implies that $s' > \frac{n}{2}$. According to Lemma 2.2-(iii), the solution of $\mathcal{A}y + ay = |\varphi_h|^{s-1}\mathrm{sign}(\varphi_h)$ belongs to $H_0^1(\Omega)\cap C(\bar\Omega)$ and satisfies $\|y\|_{C(\bar\Omega)} \leq C_{s'}\||\varphi_h|^{s-1}\mathrm{sign}(\varphi_h)\|_{L^{s'}(\Omega)} = C_{s'}\|\varphi_h\|^{s-1}_{L^s(\Omega)}$, where $C_{s'}$ is independent of a and h. Using these facts we infer

$$\begin{aligned} \|\varphi_h\|^s_{L^s(\Omega)} &= \int_\Omega |\varphi_h|^s\,\mathrm{d}x = \langle \mathcal{A}y + ay, \varphi_h\rangle = \langle \mathcal{A}^*\varphi_h + a\varphi_h, y\rangle \\ &= \int_\Omega hy\,\mathrm{d}x \leq \|h\|_{L^1(\Omega)}\|y\|_{C(\bar\Omega)} \leq C_{s'}\|h\|_{L^1(\Omega)}\|\varphi_h\|^{s-1}_{L^s(\Omega)}. \end{aligned} \qquad \square$$

This proves (2.3) for φ_h.

2.2. Analysis of the semilinear equation. In this subsection, we formulate some results concerning the semilinear equation (1.1). For this purpose we make the following assumptions on the nonlinear term of the equation.

ASSUMPTION 2.4. *We assume that $f : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function* *of class C^2 with respect to the second variable satisfying:*

$$f(\cdot,0) \in L^r(\Omega) \text{ with } r > \frac{n}{2} \text{ and } \frac{\partial f}{\partial y}(x,y) \geq 0 \ \forall y \in \mathbb{R}, \tag{2.4}$$

$$\forall M > 0 \ \exists C_{f,M} > 0 \text{ such that } \left|\frac{\partial f}{\partial y}(x,y)\right| + \left|\frac{\partial^2 f}{\partial y^2}(x,y)\right| \leq C_{f,M} \ \forall |y| \leq M, \tag{2.5}$$

$$\begin{cases} \forall M > 0 \text{ and } \forall \varepsilon > 0 \ \exists \delta > 0 \text{ such that} \\ \left|\frac{\partial^2 f}{\partial y^2}(x,y_2) - \frac{\partial^2 f}{\partial y^2}(x,y_1)\right| < \varepsilon \text{ if } |y_1|, |y_2| \leq M \text{ and } |y_2 - y_1| \leq \delta, \end{cases} \tag{2.6}$$

for almost every $x \in \Omega$.

THEOREM 2.5. *Let Assumptions 2.1 and 2.4 hold. If u belongs to $L^r(\Omega)$ for some $r > n/2$, then there exists a unique solution $y_u \in H_0^1(\Omega) \cap C(\bar{\Omega})$ of (1.1). Moreover, there exists a constant $K_{f,r}$ independent of u such that*

$$\|y_u\|_{H_0^1(\Omega)} + \|y_u\|_{C(\bar{\Omega})} \leq K_{f,r}\big(\|u\|_{L^r(\Omega)} + \|f(\cdot,0)\|_{L^r(\Omega)} + 1\big). \tag{2.7}$$

Further, if $\{u_k\}_{k=1}^\infty$ is a sequence converging weakly to u in $L^r(\Omega)$, then $y_{u_k} \to y_u$ strongly in $H_0^1(\Omega) \cap C(\bar{\Omega})$.

The reader is referred to [6] for the proof of this result. As a consequence of (2.7) we get

$$\exists K_U > 0 \text{ such that } \|y_u\|_{H_0^1(\Omega)} + \|y_u\|_{C(\bar{\Omega})} \leq K_U \quad \forall u \in \mathcal{U}_{ad}. \tag{2.8}$$

For each $r > n/2$, we define the map $G_r : L^r(\Omega) \to H_0^1(\Omega) \cap C(\bar{\Omega})$ by $G_r(u) = y_u$.

THEOREM 2.6. *Let Assumptions 2.1 and 2.4 hold. For every $r > \frac{n}{2}$ the map G_r is of class C^2, and the first and second derivatives at $u \in L^r(\Omega)$ in the directions $v, v_1, v_2 \in L^r(\Omega)$, denoted by $z_{u,v} = G_r'(u)v$ and $z_{u,v_1,v_2} = G_r''(u)(v_1,v_2)$, are the solutions of the equations*

$$\mathcal{A}z + \frac{\partial f}{\partial y}(x,y_u)z = v, \tag{2.9}$$

$$\mathcal{A}z + \frac{\partial f}{\partial y}(x,y_u)z = -\frac{\partial^2 f}{\partial y^2}(x,y_u)z_{u,v_1}z_{u,v_2}, \tag{2.10}$$

respectively.

The proof of this theorem is an easy application of the implicit function theorem; see [6].

LEMMA 2.7. *The following statements are fulfilled.*

(i) Suppose that $r > \frac{n}{2}$ and $s \in [1, \frac{n}{n-2})$. Then, there exist constants K_r depending on r and M_s depending on s such that for every $u, \bar{u} \in \mathcal{U}_{ad}$

$$\|y_u - y_{\bar{u}} - z_{\bar{u},u-\bar{u}}\|_{C(\bar{\Omega})} \leq K_r \|y_u - y_{\bar{u}}\|^2_{L^{2r}(\Omega)}, \tag{2.11}$$

$$\|y_u - y_{\bar{u}} - z_{\bar{u},u-\bar{u}}\|_{L^s(\Omega)} \leq M_s \|y_u - y_{\bar{u}}\|^2_{L^2(\Omega)}. \tag{2.12}$$

(ii) Taking $C_X = K_2\sqrt{|\Omega|}$ if $X = C(\bar{\Omega})$ and $C_X = M_2$ if $X = L^2(\Omega)$, the following inequality holds

$$\|z_{u,v} - z_{\bar{u},v}\|_X \leq C_X \|y_u - y_{\bar{u}}\|_X \|z_{\bar{u},v}\|_X \ \forall u, \bar{u} \in \mathcal{U}_{ad} \text{ and } \forall v \in L^2(\Omega). \tag{2.13}$$

(iii) There exists $\varepsilon > 0$ such that for all $\bar{u}, u \in \mathcal{U}_{ad}$ with $\|y_u - y_{\bar{u}}\|_{C(\bar{\Omega})} \leq \varepsilon$ the following inequalities are satisfied

$$\frac{1}{2}\|y_u - y_{\bar{u}}\|_X \leq \|z_{\bar{u},u-\bar{u}}\|_X \leq \frac{3}{2}\|y_u - y_{\bar{u}}\|_X, \tag{2.14}$$

$$\frac{1}{2}\|z_{\bar{u},v}\|_X \leq \|z_{u,v}\|_X \leq \frac{3}{2}\|z_{\bar{u},v}\|_X \quad \forall v \in L^2(\Omega). \tag{2.15}$$

Proof. Let us set $\phi = y_u - y_{\bar{u}} - z_{\bar{u},u-\bar{u}} \in H_0^1(\Omega) \cap C(\bar{\Omega})$. From the equations satisfied by the three functions and using the mean value theorem we get

$$\mathcal{A}\phi + \frac{\partial f}{\partial y}(x, y_{\bar{u}})\phi = \Big[\frac{\partial f}{\partial y}(x, y_{\bar{u}}) - \frac{\partial f}{\partial y}(x, y_\theta)\Big](y_u - y_{\bar{u}}),$$

where $y_\theta(x) = y_{\bar{u}}(x) + \theta(x)(y_u(x) - y_{\bar{u}}(x))$ with $\theta : \Omega \longrightarrow [0,1]$ measurable. Using again the mean value theorem we deduce

$$\mathcal{A}\phi + \frac{\partial f}{\partial y}(x, y_{\bar{u}})\phi = -\theta\frac{\partial^2 f}{\partial y^2}(x, y_\vartheta)(y_u - y_{\bar{u}})^2$$

with $y_\vartheta(x) = y_{\bar{u}}(x) + \vartheta(x)(y_\theta(x) - y_{\bar{u}}(x))$ and $\vartheta : \Omega \longrightarrow [0,1]$ measurable. By Lemma 2.2-(iii) and taking into account (2.5) and (2.8) we infer the existence of C_r independent of $u, \bar{u} \in \mathcal{U}_{ad}$ such that

$$\|\phi\|_{C(\bar{\Omega})} \leq C_r C_{f,K_U}\|(y_u - y_{\bar{u}})^2\|_{L^r(\Omega)} = C_r C_{f,K_U}\|y_u - y_{\bar{u}}\|^2_{L^{2r}(\Omega)},$$

which proves (2.11) with $K_r = C_r C_{f,K_U}$. To prove (2.12) we use Lemma 2.3 to obtain

$$\|\phi\|_{L^s(\Omega)} \leq C_{s'} C_{f,K_U}\|(y_u - y_{\bar{u}})^2\|_{L^1(\Omega)} = C_{s'} C_{f,K_U}\|y_u - y_{\bar{u}}\|^2_{L^2(\Omega)}.$$

Taking $M_s = C_{s'} C_{f,K_U}$, (2.12) follows.

Now we prove (2.13) for $X = C(\bar{\Omega})$. Setting $\psi = z_{u,v} - z_{\bar{u},v}$ and subtracting the corresponding equations we infer with the mean value theorem

$$\mathcal{A}\psi + \frac{\partial f}{\partial y}(x, y_u)\psi = \Big[\frac{\partial f}{\partial y}(x, y_{\bar{u}}) - \frac{\partial f}{\partial y}(x, y_u)\Big]z_{\bar{u},v} = \frac{\partial^2 f}{\partial y^2}(x, y_\theta)(y_{\bar{u}} - y_u)z_{\bar{u},v}.$$

Taking $r = 2$ in (2.2) and using (2.5) and (2.8) it follows from the above equation

$$\|\psi\|_{C(\bar{\Omega})} \leq C_2 C_{f,K_U}\|(y_{\bar{u}} - y_u)z_{\bar{u},v}\|_{L^2(\Omega)} \leq K_2\sqrt{|\Omega|}\|y_u - y_{\bar{u}}\|_{C(\bar{\Omega})}\|z_{\bar{u},v}\|_{C(\bar{\Omega})},$$

which proves (2.13) for $X = C(\bar{\Omega})$. The proof for $X = L^2(\Omega)$ is analogous, we use the estimate (2.3) for $s = 2$ instead of (2.2).

To prove (2.14) for $X = C(\bar{\Omega})$ we use (2.11) with $r = 2$ to get

$$\begin{aligned}\|y_u - y_{\bar{u}}\|_{C(\bar{\Omega})} &\leq \|\phi\|_{C(\bar{\Omega})} + \|z_{\bar{u},u-\bar{u}}\|_{C(\bar{\Omega})} \leq K_2\|y_u - y_{\bar{u}}\|^2_{L^4(\Omega)} + \|z_{\bar{u},u-\bar{u}}\|_{C(\bar{\Omega})}\\ &\leq K_2\sqrt{|\Omega|}\|y_u - y_{\bar{u}}\|^2_{C(\bar{\Omega})} + \|z_{\bar{u},u-\bar{u}}\|_{C(\bar{\Omega})}.\end{aligned}$$

Choosing $\varepsilon_1 = [2K_2\sqrt{|\Omega|}]^{-1}$ the first inequality of (2.14) follows if $\|y_u - y_{\bar{u}}\|_{C(\bar{\Omega})} < \varepsilon_1$. To deal with the case $X = L^2(\Omega)$ we use (2.12) with $s = 2$ and obtain

$$\begin{aligned}\|y_u - y_{\bar{u}}\|_{L^2(\Omega)} &\leq \|\phi\|_{L^2(\Omega)} + \|z_{\bar{u},u-\bar{u}}\|_{L^2(\Omega)} \leq M_2\|y_u - y_{\bar{u}}\|^2_{L^2(\Omega)} + \|z_{\bar{u},u-\bar{u}}\|_{L^2(\Omega)}\\ &\leq M_2\sqrt{|\Omega|}\|y_u - y_{\bar{u}}\|_{C(\bar{\Omega})}\|y_u - y_{\bar{u}}\|_{L^2(\Omega)} + \|z_{\bar{u},u-\bar{u}}\|_{L^2(\Omega)}.\end{aligned}$$

Hence, taking $\varepsilon_2 = [2M_2\sqrt{|\Omega|}]^{-1}$ we obtain the first inequality of (2.14) with $X = L^2(\Omega)$ if $\|y_u - y_{\bar u}\|_{C(\bar\Omega)} < \varepsilon_2$.

To prove the second inequality of (2.14) for $X = C(\bar\Omega)$, we proceed as follows

$$\|z_{\bar u,u-\bar u}\|_{C(\bar\Omega)} \le \|\phi\|_{C(\bar\Omega)} + \|y_u - y_{\bar u}\|_{C(\bar\Omega)} \le K_2\sqrt{|\Omega|}\|y_u - y_{\bar u}\|^2_{C(\bar\Omega)} + \|y_u - y_{\bar u}\|_{C(\bar\Omega)}$$
$$\le \frac{3}{2}\|y_u - y_{\bar u}\|_{C(\bar\Omega)} \quad \text{if } \|y_u - y_{\bar u}\|_{C(\bar\Omega)} < \varepsilon_1.$$

Similarly the second inequality of (2.14) follows if $X = L^2(\Omega)$ with ε_2 replacing ε_1.

Finally, we prove (2.15). Using (2.13) we obtain

$$\|z_{u,v}\|_X \le \|z_{u,v} - z_{\bar u,v}\|_X + \|z_{\bar u,v}\| \le C_X\|y_u - y_{\bar u}\|_X\|z_{\bar u,v}\|_X + \|z_{\bar u,v}\|_X,$$
$$\|z_{\bar u,v}\|_X \le \|z_{u,v} - z_{\bar u,v}\|_X + \|z_{u,v}\| \le C_X\|y_u - y_{\bar u}\|_X\|z_{\bar u,v}\|_X + \|z_{u,v}\|_X. \quad \square$$

Therefore, selecting $\varepsilon = \frac{1}{2C_X}$, then (2.15) follows if $\|y_u - y_{\bar u}\|_{C(\bar\Omega)} \le \varepsilon$.

3. The Control Problem. In this section, we make assumptions on the objective functional J so that (P) has at least one solution and the first and second order conditions for local optimality can be established. Since the problem is not convex, we will consider not only global minimizers, but also local minimizers. Throughout this paper, we will say that $\bar u$ is local minimizer of (P) if $\bar u \in \mathcal{U}_{ad}$ and there exists a ball $B_\rho(\bar u) \subset L^2(\Omega)$ such that $J(\bar u) \le J(u)$ for every $u \in \mathcal{U}_{ad} \cap B_\rho(\bar u)$. We will also say that $\bar u$ is a strong local minimizer of (P) if $\bar u \in \mathcal{U}_{ad}$ and there exists $\varepsilon > 0$ such that $J(\bar u) \le J(u)$ for every $u \in \mathcal{U}_{ad}$ with $\|y_u - y_{\bar u}\|_{C(\bar\Omega)} < \varepsilon$. If the previous inequalities are strict whenever $u \ne \bar u$, then we say that $\bar u$ is a strict (strong) local minimizer. As far as we know, the notion of strong local minimizers in the framework of control of partial differential equations was introduced for the first time in [1]; see also [2].

We make the following assumptions on L.

Assumption 3.1. *The function $L : \Omega \times \mathbb{R}^2 \longrightarrow \mathbb{R}$ is Carathéodory and of class C^2 with respect to the second variable. In addition, we assume that*

$$L(x,y,u) = L_0(x,y) + g(x)u \quad \text{with} \quad L_0(\cdot,0) \in L^1(\Omega) \quad \text{and} \quad g \in L^\infty(\Omega), \tag{3.1}$$

$$\begin{cases} \forall M > 0\ \exists \psi_M \in L^2(\Omega) \text{ and } C_{L,M} > 0 \text{ such that} \\ \Big|\dfrac{\partial L}{\partial y}(x,y,u)\Big| \le \psi_M(x) \quad \text{and} \quad \Big|\dfrac{\partial^2 L}{\partial y^2}(x,y,u)\Big| \le C_{L,M}\ \forall |y| \le M, \end{cases} \tag{3.2}$$

$$\begin{cases} \forall M > 0 \text{ and } \forall \varepsilon > 0\ \exists \delta > 0 \text{ such that} \\ \Big|\dfrac{\partial^2 L}{\partial y^2}(x,y_2,u) - \dfrac{\partial^2 L}{\partial y^2}(x,y_1,u)\Big| < \varepsilon \text{ if } |y_1|,|y_2| \le M,\ |y_2 - y_1| \le \delta, \end{cases} \tag{3.3}$$

for almost every $x \in \Omega$.

Using Theorem 2.5, the assumptions on L, and the boundedness of $\mathcal{U}_{ad}$ in $L^\infty(\Omega)$, the existence of at least one solution of (P) follows. Indeed, if we take a minimizing sequence $\{u_k\}_{k=1}^\infty$, we can assume that $u_k \stackrel{*}{\rightharpoonup} \bar u$ in $L^\infty(\Omega)$. Then Theorem 2.5 implies that $y_{u_k} \to y_{\bar u}$ strongly in $H_0^1(\Omega) \cap C(\bar\Omega)$. Further, using (2.8) and (3.2) with $M = K_U$ we infer with the mean value theorem

$$|L_0(x, y_{u_k}(x))| \le |L_0(x,0)| + \psi_{K_U}(x)K_U.$$

Then we can apply Lebesgue's dominated convergence theorem to pass to the limit in the objective functional and to obtain $J(u_k) \to J(\bar u)$.

In order to derive the first order optimality conditions satisfied by a local minimizer we address the issue of the differentiability of the objective functional J.

THEOREM 3.2. *Suppose that* $r > \frac{n}{2}$. *Then, the functional* $J : L^r(\Omega) \longrightarrow \mathbb{R}$ *is of class* C^2. *Moreover, given* $u, v, v_1, v_2 \in L^r(\Omega)$ *we have*

$$J'(u)v = \int_\Omega (\varphi_u + g)v \,\mathrm{d}x, \tag{3.4}$$

$$J''(u)(v_1, v_2) = \int_\Omega \Big[\frac{\partial^2 L}{\partial y^2}(x, y_u, u) - \varphi_u \frac{\partial^2 f}{\partial y^2}(x, y_u)\Big] z_{u,v_1} z_{u,v_2} \,\mathrm{d}x, \tag{3.5}$$

where $\varphi_u \in H_0^1(\Omega) \cap C(\bar\Omega)$ *is the unique solution of the adjoint equation*

$$\begin{cases} \mathcal{A}^*\varphi + \dfrac{\partial f}{\partial y}(x, y_u)\varphi = \dfrac{\partial L}{\partial y}(x, y_u, u) \text{ in } \Omega, \\ \varphi = 0 \text{ on } \Gamma. \end{cases} \tag{3.6}$$

This is a straightforward consequence of Theorem 2.6, Assumption 3.1, and the chain rule. The only critical issue is the existence, uniqueness, and regularity of φ_u. But this an immediate consequence of Lemma 2.2-(iii) that, as already mentioned, applies to the operator $\mathcal{A}^*$ as well. From this theorem, the optimality conditions follows in the classical way.

THEOREM 3.3. *Let* $\bar u$ *be a (strong or not) local minimizer of* (P), *then there exist two unique elements* $\bar y, \bar\varphi \in H_0^1(\Omega) \cap C(\bar\Omega)$ *such that*

$$\begin{cases} \mathcal{A}\bar y + f(x, \bar y) = \bar u \text{ in } \Omega, \\ \bar y = 0 \text{ on } \Gamma, \end{cases} \tag{3.7}$$

$$\begin{cases} \mathcal{A}^*\bar\varphi + \dfrac{\partial f}{\partial y}(x, \bar y)\bar\varphi = \dfrac{\partial L}{\partial y}(x, \bar y, \bar u) \text{ in } \Omega, \\ \bar\varphi = 0 \text{ on } \Gamma, \end{cases} \tag{3.8}$$

$$\int_\Omega (\bar\varphi + g)(u - \bar u)\, dx \geq 0 \quad \forall u \in \mathcal{U}_{ad}. \tag{3.9}$$

The derivation of sufficient second order conditions for local optimality is more delicate. First we introduce the cone of critical directions on which we formulate the necessary second order conditions for optimality: if $\bar u \in \mathcal{U}_{ad}$ is a local minimizer of (P) we define

$$C_{\bar u} = \{v \in L^2(\Omega) : J'(\bar u)v = 0 \text{ and } v \text{ satisfies the sign conditions (3.10)}\},$$

$$v(x)\begin{cases} \geq 0 & \text{if } \bar u(x) = u_a, \\ \leq 0 & \text{if } \bar u(x) = u_b. \end{cases} \tag{3.10}$$

As usual, from (3.9) we deduce that $(\bar\varphi + g)(x)v(x) \geq 0$ for almost all $x \in \Omega$ if $v \in L^2(\Omega)$ satisfies (3.10). Therefore, the condition $J'(\bar u)v = 0$ for v satisfying (3.10) is only possible if $v(x) = 0$ for almost every $x \in \Omega$ such that $(\bar\varphi + g)(x) \neq 0$. Therefore, $C_{\bar u}$ can be written

$$C_{\bar u} = \{v \in L^2(\Omega) : \text{satisfying (3.10) and } v(x) = 0 \text{ if } |(\bar\varphi + g)(x)| > 0\}.$$

It is well known that every local minimizer $\bar{u}$ satisfies the second order necessary optimality condition: $J''(\bar{u})v^2 \geq 0$ for all $v \in C_{\bar{u}}$; see, for instance, [8]. However, based on $C_{\bar{u}}$ it is not possible to get sufficient second order conditions for local optimality. The reader is referred to [12] for a counterexample. A procedure suggested by several authors consists in extending the cone of critical directions $C_{\bar{u}}$; see [10, 11, 18, 19]. Two possible extensions of $C_{\bar{u}}$ seem natural after the above comments: for $\tau > 0$ we define the extended cones

$$D_{\bar{u}}^{\tau} = \{v \in L^2(\Omega) : \text{satisfying (3.10) and } v(x) = 0 \text{ if } |(\bar{\varphi} + g)(x)| > \tau\},$$
$$G_{\bar{u}}^{\tau} = \{v \in L^2(\Omega) : \text{satisfying (3.10) and } J'(\bar{u})v \leq \tau \|z_v\|_{L^1(\Omega)}\}.$$

On any of these cones we can formulate sufficient second order conditions for local optimality. Obviously, both are extensions of $C_{\bar{u}}$. In [3], the authors introduced the cone $C_{\bar{u}}^{\tau} = D_{\bar{u}}^{\tau} \cap G_{\bar{u}}^{\tau}$, which is also an extension of $C_{\bar{u}}$. They proved that the first order optimality conditions (3.7)–(3.9) along with the condition

$$\exists \delta > 0 \text{ such that } J''(\bar{u})v^2 \geq \delta \|z_v\|_{L^2(\Omega)}^2 \quad \forall v \in C_{\bar{u}}^{\tau} \tag{3.11}$$

imply the existence of $\kappa > 0$ and $\varepsilon > 0$ such that

$$J(\bar{u}) + \frac{\kappa}{2}\|y_u - \bar{y}\|_{L^2(\Omega)}^2 \leq J(u) \;\; \forall u \in \mathcal{U}_{ad} \text{ such that } \|y_u - \bar{y}\|_{C(\bar{\Omega})} < \varepsilon. \tag{3.12}$$

Actually, the proof of [3] was carried out for a parabolic control problem with $g = 0$. However, the same proof works for the elliptic case and $g \neq 0$. Here, we formulate a new assumption leading to the same result (3.12) as (3.11) does.

ASSUMPTION 3.4. *There exist numbers $\alpha > 0$ and $\gamma > 0$ such that*

$$J'(\bar{u})(u-\bar{u}) + J''(\bar{u})(u-\bar{u})^2 \geq \gamma \|z_{\bar{u},u-\bar{u}}\|_{L^2(\Omega)}^2 \; \forall u \in \mathcal{U}_{ad} \text{ with } \|y_u - \bar{y}\|_{C(\bar{\Omega})} < \alpha. \tag{3.13}$$

It was proved in [4] that (3.11) implies (3.13). Therefore, (3.13) appears as a weaker assumption. However, the next theorem proves that it is sufficient to imply (3.12).

THEOREM 3.5. *Let $\bar{u} \in \mathcal{U}_{ad}$ satisfy the optimality conditions (3.7)–(3.9) and Assumption 3.4. Then, there exists $\varepsilon > 0$ and $\kappa > 0$ such that (3.12) holds.*

Before proving this theorem we establish some lemmas.

LEMMA 3.6. *Let $\bar{u} \in \mathcal{U}_{ad}$ be fixed with associated state $\bar{y}$. Then, the following inequality holds for all $\theta \in [0, 1]$ and $u \in \mathcal{U}_{ad}$*

$$\|y_{\bar{u}+\theta(u-\bar{u})} - \bar{y}\|_{C(\bar{\Omega})} \leq (C_2 C_{f,K_U} \sqrt{|\Omega|} \|y_u - \bar{y}\|_{C(\bar{\Omega})} + 1) \|y_u - \bar{y}\|_{C(\bar{\Omega})}, \tag{3.14}$$

where C_2 is the constant of (2.2) with $r = 2$ and C_{f,K_U} is the one deduced from (2.5) and (2.8).

Proof. The proof of this lemma is based on the analogous result for parabolic control problems established in [5]. We take $\theta \in [0, 1]$ and $u \in \mathcal{U}_{ad}$. We set $\phi = y_{\bar{u}+\theta(u-\bar{u})} - [\bar{y} + \theta(y_u - \bar{y})]$. Then, we have

$$\mathcal{A}\phi + f(x, y_{\bar{u}+\theta(u-\bar{u})}) - [f(x, \bar{y}) + \theta(f(x, y_u) - f(x, \bar{y}))] = 0.$$

Applying the mean value theorem, we obtain measurable functions $\theta_i : \Omega \longrightarrow [0,1]$, $i = 1, 2$, such that

$$f(x, y_{\bar u+\theta(u-\bar u)}) - f(x,\bar y) = \frac{\partial f}{\partial y}(x, y_1)(y_{\bar u+\theta(u-\bar u)} - \bar y) \text{ and } y_1 = \bar y + \theta_1(y_{\bar u+\theta(u-\bar u)} - \bar y),$$

$$f(x, y_u) - f(x, \bar y) = \frac{\partial f}{\partial y}(x, y_2)(y_u - \bar y) \text{ with } y_2 = \bar y + \theta_2(y_u - \bar y).$$

Inserting these identities in the above partial differential equation we infer

$$\mathcal{A}\phi + \frac{\partial f}{\partial y}(x, y_1)(y_{\bar u+\theta(u-\bar u)} - \bar y) - \theta\frac{\partial f}{\partial y}(x, y_2)(y_u - \bar y) = 0.$$

Noting that $y_{\bar u+\theta(u-\bar u)} - \bar y = \phi + \theta(y_u - \bar y)$, the above equality and a new application of the mean value theorem lead to

$$\mathcal{A}\phi + \frac{\partial f}{\partial y}(x, y_1)\phi = \theta\Big[\frac{\partial f}{\partial y}(x, y_2) - \frac{\partial f}{\partial y}(x, y_1)\Big](y_u - \bar y) = \theta\frac{\partial^2 f}{\partial y^2}(x, y_3)(y_u - \bar y)^2,$$

where $y_3 = y_1 + \theta_3(y_2 - y_1)$. Using (2.2) with $r = 2$, (2.5), and (2.8) we infer

$$\|\phi\|_{C(\bar\Omega)} \le C_2 C_{f,K_U}\|(y_u - \bar y)^2\|_{L^2(\Omega)} \le C_2 C_{f,K_U}\sqrt{|\Omega|}\|y_u - \bar y\|^2_{C(\bar\Omega)}.$$

This implies

$$\begin{aligned}&\|y_{\bar u+\theta(u-\bar u)} - \bar y\|_{C(\bar\Omega)} = \|\phi + \theta(y_u - \bar y)\|_{C(\bar\Omega)}\\ &\le (C_2 C_{f,K_U}\sqrt{|\Omega|}\|y_u - \bar y\|_{C(\bar\Omega)} + 1)\|y_u - \bar y\|_{C(\bar\Omega)}.\end{aligned}$$ □

Lemma 3.7. *There exists a constant $M_U > 0$ such that*

$$\|\varphi_u\|_{C(\bar\Omega)} \le M_U \quad \forall u \in \mathcal{U}_{ad}. \tag{3.15}$$

Moreover, given $\bar u \in \mathcal{U}_{ad}$ with associated state $\bar y$ and adjoint state $\bar\varphi$, we have

$$\|\varphi_{\bar u+\theta(u-\bar u)} - \bar\varphi\|_{C(\bar\Omega)} \le C\|y_u - \bar y\|_{C(\bar\Omega)} \quad \forall\theta \in [0,1] \text{ and } \forall u \in \mathcal{U}_{ad}, \tag{3.16}$$

where C depends only on f, L, $\mathcal{U}_{ad}$, and Ω.

Proof. For the proof of (3.15) we use (2.2) with $r = 2$, (2.8), and (3.2) as follows

$$\|\varphi_u\|_{C(\bar\Omega)} \le C_2\Big\|\frac{\partial L}{\partial y}(x, y_u, u)\Big\|_{L^2(\Omega)} \le M_U = C_2\|\psi_{K_U}\|_{L^2(\Omega)}.$$

Let us prove (3.16). Given $u \in \mathcal{U}$ and $\theta \in [0,1]$ let us denote $u_\theta = \bar u + \theta(u - \bar u)$, $y_\theta = y_{u_\theta}$, and $\varphi_\theta = \varphi_{u_\theta}$. Subtracting the equations satisfied by φ_θ and $\bar\varphi$ we get with the mean value theorem

$$\begin{aligned}&\mathcal{A}^*(\varphi_\theta - \bar\varphi) + \frac{\partial f}{\partial y}(x, \bar y)(\varphi_\theta - \bar\varphi) = \frac{\partial L}{\partial y}(x, y_\theta, u_\theta) - \frac{\partial L}{\partial y}(x, \bar y, \bar u)\\ &+ \Big[\frac{\partial f}{\partial y}(x, \bar y) - \frac{\partial f}{\partial y}(x, y_\theta)\Big]\varphi_\theta = \Big[\frac{\partial^2 L}{\partial y^2}(x, y_\vartheta, u_\vartheta) - \varphi_\theta\frac{\partial^2 f}{\partial y^2}(x, y_\vartheta)\Big](y_\theta - \bar y),\end{aligned}$$

where $y_\vartheta = \bar y + \vartheta(y_\theta - \bar y)$ for some measurable function $\vartheta : \Omega \longrightarrow [0,1]$. Now, we apply (2.2) with $r = 2$, (2.8), (3.15), (2.5), and (3.2) to get from the above equation

$$\|\varphi_\theta - \bar\varphi\|_{C(\bar\Omega)} \le C_2(C_{L,K_U} + M_U C_{f,K_U})\sqrt{|\Omega|}\|y_\theta - \bar y\|_{C(\bar\Omega)}.$$ □

Then, (3.16) follows from Lemma 3.6.

LEMMA 3.8. *For every* $\rho > 0$ *there exists* $\varepsilon > 0$ *such that if* $u \in \mathcal{U}_{ad}$ *and* $\|y_u - \bar{y}\|_{C(\bar{\Omega})} < \varepsilon$ *then*

$$|[J''(\bar{u} + \theta(u - \bar{u})) - J''(\bar{u})]v^2| < \rho\|z_{\bar{u},v}\|^2_{L^2(\Omega)} \quad \forall v \in L^2(\Omega) \text{ and } \forall \theta \in [0,1]. \tag{3.17}$$

Proof. First, let us denote u_θ, y_θ, and φ_θ as in the proof of Lemma 3.7. From (3.5) we get

$$\begin{aligned}
&|[J''(\bar{u} + \theta(u - \bar{u})) - J''(\bar{u})]v^2| \le \int_\Omega \Big| \Big[\frac{\partial^2 L}{\partial y^2}(x, y_\theta, u_\theta) - \frac{\partial^2 L}{\partial y^2}(x, \bar{y}, \bar{u})\Big] z^2_{u_\theta,v}\Big| \,\mathrm{d}x \\
&+ \int_\Omega \Big|(\varphi_\theta - \bar{\varphi})\frac{\partial^2 f}{\partial y^2}(x, y_\theta) z^2_{u_\theta,v}\Big| \,\mathrm{d}x + \int_\Omega \Big|\bar{\varphi}\Big[\frac{\partial^2 f}{\partial y^2}(x, y_\theta) - \frac{\partial^2 f}{\partial y^2}(x, \bar{y})\Big] z^2_{u_\theta,v}\Big| \,\mathrm{d}x \\
&+ \int_\Omega \Big|\Big[\frac{\partial^2 L}{\partial y^2}(x, \bar{y}, \bar{u}) - \bar{\varphi}\frac{\partial^2 f}{\partial y^2}(x, \bar{y})\Big](z^2_{u_\theta,v} - z^2_{\bar{u},v})\Big| \,\mathrm{d}x \\
&= I_1 + I_2 + I_3 + I_4.
\end{aligned}$$

Let us estimate the terms I_i. For I_1 we deduce from (3.3), (2.15), and (3.14) that for every $\rho > 0$ there exists $\varepsilon > 0$ such that $I_1 \le \rho\|z_{\bar{u},v}\|^2_{L^2(\Omega)}$ if $\|y_u - \bar{y}\|_{C(\bar{\Omega})} < \varepsilon$. The same estimate can be deduced for I_2 using (2.5), (2.8), (2.15), and (3.16). The estimate for I_3 follows from (2.6), (2.8), (2.15), (3.14), and (3.15). Finally, we estimate I_4 by using (2.5), (2.8), (2.13), (2.15), (3.2), (3.14), and (3.15) to infer that

$$\begin{aligned}
I_4 &\le (C_{L,K_U} + M_U C_{f,K_U})\|z_{u_\theta,v} + z_{\bar{u},v}\|_{L^2(\Omega)}\|z_{u_\theta,v} - z_{\bar{u},v}\|_{L^2(\Omega)} \\
&\le (C_{L,K_U} + M_U C_{f,K_U})\frac{5}{2}\|z_{\bar{u},v}\|_{L^2(\Omega)} C_{L^2(\Omega)}\|y_\theta - \bar{y}\|_{C(\bar{\Omega})}\|z_{\bar{u},v}\|_{L^2(\Omega)} \\
&\le \rho\|z_{\bar{u},v}\|^2_{L^2(\Omega)} \quad \text{if } \|y_u - \bar{y}\|_{C(\bar{\Omega})} < \varepsilon.
\end{aligned}$$
∎

Hence, (3.17) is a straightforward consequence of the above estimates.

Proof of Theorem 3.5. Let us take $u \in \mathcal{U}_{ad}$ with $\|y_u - \bar{y}\|_{C(\bar{\Omega})} < \alpha$. By performing a Taylor expansion and using that $J'(\bar{u})(u - \bar{u}) \ge 0$ we obtain

$$\begin{aligned}
J(u) &= J(\bar{u}) + J'(\bar{u})(u - \bar{u}) + \frac{1}{2}J''(u_\theta)(u - \bar{u})^2 \\
&\ge J(\bar{u}) + \frac{1}{2}[J'(\bar{u})(u - \bar{u}) + J''(\bar{u})(u - \bar{u})^2] + \frac{1}{2}[J''(u_\theta) - J''(\bar{u})](u - \bar{u})^2 \\
&\ge J(\bar{u}) + \frac{\delta}{2}\|z_{\bar{u},u-\bar{u}}\|^2_{L^2(\Omega)} - \frac{1}{2}|[J''(u_\theta) - J''(\bar{u})](u - \bar{u})^2|.
\end{aligned}$$

Lemma 3.8 implies the existence of $\varepsilon \in (0, \alpha]$ such that $|[J''(u_\theta) - J''(\bar{u})](u - \bar{u})^2| < \frac{\delta}{2}\|z_{\bar{u},u-\bar{u}}\|^2_{L^2(\Omega)}$ for every $u \in \mathcal{U}_{ad}$ with $\|y_u - \bar{y}\|_{C(\bar{\Omega})} < \varepsilon$. Inserting this estimate in the above expression and taking ε still smaller if necessary, we can apply (2.14) to deduce

$$J(u) \ge J(\bar{u}) + \frac{\delta}{4}\|z_{\bar{u},u-\bar{u}}\|^2_{L^2(\Omega)} \ge J(\bar{u}) + \frac{\delta}{16}\|y_u - \bar{y}\|^2_{L^2(\Omega)}. \qquad \textit{None}$$

This inequality yields (3.12) with $\kappa = \frac{\delta}{8}$.

4. Stability of the states. In this section, we consider the following perturbations of the control problem (P)

$$(\mathrm{P}) \quad \min_{u\in\mathcal{U}_{ad}} J_\varepsilon(u) := \int_\Omega [L(x, y_u^\varepsilon(x), u(x)) + \eta_\varepsilon(x) y_u^\varepsilon(x)]\,\mathrm{d}x,$$

where y_u^ε is the solution of the equation

$$\left\{ \begin{array}{l} -\mathrm{div}\big(A(x)\nabla y\big) + b(x)\cdot\nabla y + f(x,y) = u + \xi_\varepsilon \ \text{ in } \ \Omega, \\ y = 0 \ \text{ on } \ \Gamma. \end{array} \right. \tag{4.1}$$

Here we assume that $\{\xi_\varepsilon\}_{\varepsilon>0}$ and $\{\eta_\varepsilon\}_{\varepsilon>0}$ are bounded families in $L^2(\Omega)$ satisfying that $(\xi_\varepsilon, \eta_\varepsilon) \to (0,0)$ in $L^2(\Omega)^2$ as $\varepsilon \to 0$. As a consequence of Theorem 2.5 we get the existence and uniqueness of a solution $y_u^\varepsilon \in H_0^1(\Omega)\cap C(\bar\Omega)$ of (4.1). Moreover, using (2.7) with $r = 2$ and the boundedness of $\{\xi_\varepsilon\}_{\varepsilon>0}$ in $L^2(\Omega)$ we infer that the set $\{y_u^\varepsilon : u \in \mathcal{U}_{ad} \text{ and } \varepsilon > 0\}$ is bounded in $H_0^1(\Omega)\cap C(\bar\Omega)$. Therefore, increasing the value of K_U, if necessary, we can assume that (2.8) and the inequality

$$\|y_u^\varepsilon\|_{H_0^1(\Omega)} + \|y_u^\varepsilon\|_{C(\bar\Omega)} \le K_U \quad \forall u \in \mathcal{U}_{ad} \ \text{ and } \ \forall \varepsilon > 0 \tag{4.2}$$

hold. We will prove that the solutions of problems (P_ε) converge to the solutions of (P) in some sense to be precised below. Conversely, we will also prove that any strict strong local minimizer of (P) can be approximated by strong local minimizers of problems (P_ε). Finally, the Lipschitz stability of the optimal states with respect to the perturbations is established. We start analyzing the difference between the solutions of (1.1) and (4.1).

THEOREM 4.1. *The following inequalities hold for every $\varepsilon > 0$*

$$\|y_u^\varepsilon - y_u\|_{H_0^1(\Omega)} + \|y_u^\varepsilon - y_u\|_{C(\bar\Omega)} \le C_2\|\xi_\varepsilon\|_{L^2(\Omega)} \ \forall u \in L^2(Q), \tag{4.3}$$

$$\|z_{u,v}^\varepsilon - z_{u,v}\|_{L^2(\Omega)} \le C_2^2 C_{f,K_U}\|\xi_\varepsilon\|_{L^2(\Omega)}\|z_{u,v}\|_{L^2(\Omega)} \ \forall (u,v) \in \mathcal{U}_{ad}\times L^2(\Omega), \tag{4.4}$$

where C_2 is the constant given in (2.2) for $r = 2$, C_{f,K_U} is the constant $C_{f,M}$ of (2.5) with $M = K_U$ given in (2.8) or (4.2), and $z_{u,v}^\varepsilon$ denotes the solution of (2.9) with y_u^ε replacing y_u.

Proof. Subtracting the equations (4.1) and (1.1) and using the mean value theorem we obtain

$$\mathcal{A}(y_u^\varepsilon - y_u) + \frac{\partial f}{\partial y}(x, y_\theta)(y_u^\varepsilon - y_u) = \xi_\varepsilon.$$

Then, (2.2) implies (4.3). To prove (4.4) we subtract the equations satisfied by $z_{u,v}^\varepsilon$ and $z_{u,v}$ to obtain

$$\mathcal{A}(z_{u,v}^\varepsilon - z_{u,v}) + \frac{\partial f}{\partial y}(x, y_u^\varepsilon)(z_{u,v}^\varepsilon - z_{u,v}) = \Big[\frac{\partial f}{\partial y}(x, y_u) - \frac{\partial f}{\partial y}(x, y_u^\varepsilon)\Big] z_{u,v}.$$

Now, using (2.3) with $s = 2$, (2.5), (2.8), and (4.3) we obtain from the previous equation with the mean value theorem

$$\begin{aligned} \|z_{u,v}^\varepsilon - z_{u,v}\|_{L^2(\Omega)} &\le C_2\Big\|\Big[\frac{\partial f}{\partial y}(x, y_u) - \frac{\partial f}{\partial y}(x, y_u^\varepsilon)\Big] z_{u,v}\Big\|_{L^1(\Omega)} \\ &\le C_2 C_{f,K_U}\|(y_u^\varepsilon - y_u) z_{u,v}\|_{L^1(\Omega)} \\ &\le C_2 C_{f,K_U}\|y_u^\varepsilon - y_u\|_{L^2(\Omega)}\|z_{u,v}\|_{L^2(\Omega)} \le C_2^2 C_{f,K_U}\|\xi_\varepsilon\|_{L^2(\Omega)}\|z_{u,v}\|_{L^2(\Omega)}. \end{aligned}$$

∎

Now we analyze the convergence of problems (P_ε) to (P).

THEOREM 4.2. *Let $\{u_\varepsilon\}_{\varepsilon>0}$ be a family of solutions of problems* (P_ε)*. Any control $\bar u$ that is a weak* limit in $L^\infty(\Omega)$ of a sequence $\{u_{\varepsilon_k}\}_{k=1}^\infty$ with $\varepsilon_k \to 0$ as $k \to \infty$ is a solution of* (P)*. Moreover, the strong convergence $y^{\varepsilon_k}_{u_{\varepsilon_k}} \to y_{\bar u}$ in $H_0^1(\Omega)\cap C(\bar\Omega)$ holds.*

Proof. The existence of the sequences $\{u_{\varepsilon_k}\}_{k=1}^\infty$ converging to $\bar u$ weakly* in $L^\infty(\Omega)$ is a consequence of the boundedness of $\mathcal{U}_{ad}$ in $L^\infty(\Omega)$. From Theorem 2.5 and (4.3) we infer

$$\begin{aligned}
&\|y^{\varepsilon_k}_{u_{\varepsilon_k}} - y_{\bar u}\|_{H_0^1(\Omega)} + \|y^{\varepsilon_k}_{u_{\varepsilon_k}} - y_{\bar u}\|_{C(\bar\Omega)}\\
&\le \|y^{\varepsilon_k}_{u_{\varepsilon_k}} - y_{u_{\varepsilon_k}}\|_{H_0^1(\Omega)} + \|y^{\varepsilon_k}_{u_{\varepsilon_k}} - y_{u_{\varepsilon_k}}\|_{C(\bar\Omega)} + \|y_{u_{\varepsilon_k}} - y_{\bar u}\|_{H_0^1(\Omega)} + \|y_{u_{\varepsilon_k}} - y_{\bar u}\|_{C(\bar\Omega)}\\
&\le C_2\|\xi_\varepsilon\|_{L^2(\Omega)} + \|y_{u_{\varepsilon_k}} - y_{\bar u}\|_{H_0^1(\Omega)} + \|y_{u_{\varepsilon_k}} - y_{\bar u}\|_{C(\bar\Omega)} \to 0 \quad \text{as } k\to\infty.
\end{aligned}$$

Using this fact, the convergence $\eta_\varepsilon \to 0$ as $\varepsilon \to 0$, (3.2), the optimality of u_{ε_k} for ($\mathrm{P}_{\varepsilon_k}$), and again (4.3), we get

$$J(\bar u) = \lim_{k\to\infty} J_{\varepsilon_k}(u_{\varepsilon_k}) \le \lim_{k\to\infty} J_{\varepsilon_k}(u) = J(u) \quad \forall u \in \mathcal{U}_{ad}, \qquad \square$$

which proves that $\bar u$ is a solution of (P).

Now, we establish a kind of converse result.

THEOREM 4.3. *Let $\bar u$ be a strict strong local minimizer of* (P)*. Then, there exist $\varepsilon_0 > 0$ and a family of strong local minimizers $\{u_\varepsilon\}_{\varepsilon<\varepsilon_0}$ of problems* (P_ε) *such that $u_\varepsilon \stackrel{*}{\rightharpoonup} \bar u$ in $L^\infty(\Omega)$ and $y^\varepsilon_{u_\varepsilon} \to y_{\bar u}$ strongly in $H_0^1(\Omega)\cap C(\bar\Omega)$ as $\varepsilon \to 0$.*

Proof. Since $\bar u$ is a strict strong local minimizer of (P), there exists $\rho > 0$ such that $\bar u$ is the unique solution of the problem

$$(\mathrm{P}_\rho) \ \min_{u\in\mathcal{U}_\rho} J(u),$$

where $\mathcal{U}_\rho = \{u \in \mathcal{U}_{ad} : \|y_u - y_{\bar u}\|_{C(\bar\Omega)} \le \rho\}$. Now, for every $\varepsilon > 0$ we define the problems

$$(\mathrm{P}_{\rho,\varepsilon}) \ \min_{u\in\mathcal{U}_\rho} J_\varepsilon(u). \qquad \square$$

Using Theorem 2.5 we deduce that $\mathcal{U}_\rho$ is weakly* closed in $L^\infty(\Omega)$, hence the existence of a solution u_ε of ($\mathrm{P}_{\rho,\varepsilon}$) can be proved as we indicated for (P). Moreover, arguing as in the proof of Theorem 4.2, we deduce the existence of sequences $\{u_{\varepsilon_k}\}_{k=1}^\infty$ converging weakly* to a solution u of (P_ρ) in $L^\infty(\Omega)$ and such that $y^{\varepsilon_k}_{u_{\varepsilon_k}} \to y_u$ strongly in $H_0^1(\Omega)\cap C(\bar\Omega)$. Since $\bar u$ is the unique solution of (P_ρ), we conclude the convergence $u_\varepsilon \stackrel{*}{\rightharpoonup} \bar u$ in $L^\infty(\Omega)$ and $y^\varepsilon_{u_\varepsilon} \to y_{\bar u}$ in $H_0^1(\Omega)\cap C(\bar\Omega)$ as $\varepsilon\to 0$. Therefore, there exists $\varepsilon_0 > 0$ such that $\|y^\varepsilon_{u_\varepsilon} - y_{\bar u}\|_{C(\bar\Omega)} < \rho$ for every $\varepsilon < \varepsilon_0$. This implies that u_ε is a strong local minimizer of (P_ε) for every $\varepsilon < \varepsilon_0$, which completes the proof.

Now we establish our main theorem of this section.

THEOREM 4.4. *Let $\bar u$ be a local minimizer of* (P) *satisfying Assumption 3.4 and $\{u_\varepsilon\}_{\varepsilon<\varepsilon_0}$ a family of local solutions of problems* (P_ε) *such that $u_\varepsilon \stackrel{*}{\rightharpoonup} \bar u$ in $L^\infty(\Omega)$ as $\varepsilon \to 0$. Then, there exist $\hat\varepsilon \in (0,\varepsilon_0)$ and a constant $C > 0$ such that*

$$\|y^\varepsilon_{u_\varepsilon} - \bar y\|_{L^2(\Omega)} \le C\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big) \quad \forall \varepsilon < \hat\varepsilon, \tag{4.5}$$

where $\bar y = y_{\bar u}$.

Let us observe that Assumption 3.4 implies that $\bar u$ satisfies (3.12). Hence, $\bar u$ is a strict strong local minimizer of (P) and, consequently, Theorem 4.3 ensures the existence of a family $\{u_\varepsilon\}_{\varepsilon<\varepsilon_0}$ of strong local minimizers of problems (P$_\varepsilon$) satisfying the conditions of the above theorem. Before proving this theorem we establish the following lemma.

LEMMA 4.5. *Let $\bar u$ satisfy the assumptions of Theorem 4.4. Then, there exists $\varepsilon > 0$ such that*

$$J'(u)(u-\bar u) \ge \frac{\gamma}{2}\|z_{u,u-\bar u}\|^2_{L^2(\Omega)} \quad \forall u \in \mathcal{U}_{ad} \text{ with } \|y_u - \bar y\|_{C(\bar\Omega)} < \varepsilon, \tag{4.6}$$

where γ is given in Assumption 3.4.

Proof. We denote by $H : \Omega \times \mathbb{R}^3 \longrightarrow \mathbb{R}$ the Hamiltonian associated with the control problem (P):

$$H(x,y,\varphi,u) = L(x,y,u) + \varphi[u - f(x,y)].$$

For every $u \in \mathcal{U}_{ad}$ and $v \in L^2(\Omega)$, we define $\psi_{u,v} \in H^1_0(\Omega)\cap C(\bar\Omega)$ as the function satisfying

$$\mathcal{A}^*\psi_{u,v} + \frac{\partial f}{\partial y}(x,y_u)\psi_{u,v} = \frac{\partial^2 H}{\partial y^2}(x,y_u,\varphi_u,u)z_{u,v}.$$

We split the proof into two steps.

Step I.- Here we prove that for every $\rho > 0$ there exists $\varepsilon > 0$ such that for every $u \in \mathcal{U}_{ad}$ with $\|y_u - \bar y\|_{C(\bar\Omega)} < \varepsilon$ we have

$$\Big|\int_\Omega (\varphi_u - \bar\varphi - \psi_{\bar u,u-\bar u})(u-\bar u)\,\mathrm{d}x\Big| \le \rho\|z_{\bar u,u-\bar u}\|^2_{L^2(\Omega)}. \tag{4.7}$$

Setting $\pi = \varphi_u - \bar\varphi - \psi_{\bar u,u-\bar u}$ and subtracting their respective equations it follows with the mean value theorem

$$\begin{aligned}
\mathcal{A}^*\pi + \frac{\partial f}{\partial y}(x,\bar y)\pi &= \frac{\partial H}{\partial y}(x,y_u,\varphi_u,u) - \frac{\partial H}{\partial y}(x,\bar y,\bar\varphi,\bar u)\\
&\quad - \frac{\partial^2 H}{\partial y^2}(x,\bar y,\bar\varphi,\bar u)z_{\bar u,u-\bar u} - \frac{\partial^2 H}{\partial y\partial\varphi}(x,\bar y,\bar\varphi,\bar u)(\varphi_u-\bar\varphi)\\
&= \frac{\partial^2 H}{\partial y^2}(x,y_\theta,\varphi_\theta,u_\theta)(y_u-\bar y) - \frac{\partial^2 H}{\partial y^2}(x,\bar y,\bar\varphi,\bar u)z_{\bar u,u-\bar u}\\
&\quad + \Big[\frac{\partial^2 H}{\partial y\partial\varphi}(x,y_\theta,\varphi_\theta,u_\theta) - \frac{\partial^2 H}{\partial y\partial\varphi}(x,\bar y,\bar\varphi,\bar u)\Big](\varphi_u-\bar\varphi)\\
&= \frac{\partial^2 H}{\partial y^2}(x,y_\theta,\varphi_\theta,u_\theta)(y_u-\bar y-z_{\bar u,u-\bar u})\\
&\quad + \Big[\frac{\partial^2 H}{\partial y^2}(x,y_\theta,\varphi_\theta,u_\theta) - \frac{\partial^2 H}{\partial y^2}(x,\bar y,\bar\varphi,\bar u)\Big]z_{\bar u,u-\bar u}\\
&\quad + \Big[\frac{\partial^2 H}{\partial y\partial\varphi}(x,y_\theta,\varphi_\theta,u_\theta) - \frac{\partial^2 H}{\partial y\partial\varphi}(x,\bar y,\bar\varphi,\bar u)\Big](\varphi_u-\bar\varphi).
\end{aligned}$$

This implies

$$\begin{aligned}
\int_\Omega \pi(u-\bar u)\,\mathrm{d}x &= \int_\Omega \pi\Big(\mathcal{A}z_{\bar u,u-\bar u} + \frac{\partial f}{\partial y}(x,\bar y)z_{\bar u,u-\bar u}\Big)\,\mathrm{d}x\\
&= \int_\Omega \Big(\mathcal{A}^*\pi + \frac{\partial f}{\partial y}(x,\bar y)\pi\Big)z_{\bar u,u-\bar u}\,\mathrm{d}x\\
&= \int_\Omega \frac{\partial^2 H}{\partial y^2}(x,y_\theta,\varphi_\theta,u_\theta)(y_u-\bar y - z_{\bar u,u-\bar u})z_{\bar u,u-\bar u}\,\mathrm{d}x\\
&+ \int_\Omega \Big[\frac{\partial^2 H}{\partial y^2}(x,y_\theta,\varphi_\theta,u_\theta) - \frac{\partial^2 H}{\partial y^2}(x,\bar y,\bar\varphi,\bar u)\Big]z^2_{\bar u,u-\bar u}\,\mathrm{d}x\\
&+ \Big[\frac{\partial^2 H}{\partial y\partial\varphi}(x,y_\theta,\varphi_\theta,u_\theta) - \frac{\partial^2 H}{\partial y\partial\varphi}(x,\bar y,\bar\varphi,\bar u)\Big](\varphi_u-\bar\varphi)z_{\bar u,u-\bar u}\,\mathrm{d}x = I_1+I_2+I_3.
\end{aligned}$$

We estimate every term I_i. For the first term we use (2.5), (2.8), (2.12) with $s=2$, (2.14) with $X=L^2(\Omega)$, (3.2), and (3.15) as follows

$$\begin{aligned}
|I_1| &\le (C_{L,K_U}+M_U C_{f,K_U})\|y_u-\bar y - z_{\bar u,u-\bar u}\|_{L^2(\Omega)}\|z_{\bar u,u-\bar u}\|_{L^2(\Omega)}\\
&\le (C_{L,K_U}+M_U C_{f,K_U})M_2\|y_u-\bar y\|^2_{L^2(\Omega)}\|z_{\bar u,u-\bar u}\|_{L^2(\Omega)}\\
&\le 2(C_{L,K_U}+M_U C_{f,K_U})M_2\sqrt{|\Omega|}\varepsilon\|z_{\bar u,u-\bar u}\|^2_{L^2(\Omega)}.
\end{aligned}$$

The second term is estimated with (2.6), (2.8), (3.3), (3.14), (3.15), (3.16), leading to $|I_2|\le\rho\|z_{\bar u,u-\bar u}\|^2_{L^2(\Omega)}$ for ρ arbitrarily small if ε is taken according to ρ. Finally, for the last term we use the same inequalities as for I_2 and additionally (2.15) with $X=L^2(\Omega)$ to get

$$\begin{aligned}
|I_3| &\le \rho\|\varphi_u-\bar\varphi\|_{L^2(\Omega)}\|z_{\bar u,u-\bar u}\|_{L^2(\Omega)}\\
&\le \rho C_2(C_{L,K_U}+M_U C_{f,K_U})\sqrt{|\Omega|}\|y_u-\bar y\|_{C(\bar\Omega)}\|z_{\bar u,u-\bar u}\|_{L^2(\Omega)}\\
&\le 2C_2(C_{L,K_U}+M_U C_{f,K_U})\sqrt{|\Omega|}\rho\|z_{\bar u,u-\bar u}\|^2_{L^2(\Omega)},
\end{aligned}$$

where again ρ is arbitrarily small if ε is chosen according to it. Thus, (4.7) follows from the proved estimates.

Step II- Now, we prove (4.6). First, we observe that for every $v\in L^2(\Omega)$

$$\begin{aligned}
&\int_\Omega \psi_{\bar u,v}v\,\mathrm{d}x = \int_\Omega \psi_{\bar u,v}\Big(\mathcal{A}z_{\bar u,v} + \frac{\partial f}{\partial y}(x,\bar y)z_{\bar u,v}\Big)\,\mathrm{d}x\\
&= \int_\Omega \Big(\mathcal{A}^*\psi_{\bar u,v} + \frac{\partial f}{\partial y}(x,\bar y)\psi_{\bar u,v}\Big)z_{\bar u,v}\,\mathrm{d}x = \int_\Omega \frac{\partial^2 H}{\partial y^2}(x,\bar y,\bar\varphi,\bar u)z^2_{\bar u,v}\,\mathrm{d}x = J''(\bar u)v^2,
\end{aligned}$$

where the last inequality follows from (3.5) and the definition of the Hamiltonian. Let $\varepsilon>0$ be such that (4.7) holds with $\rho=\frac{\gamma}{2}$. Then, using Assumption 3.4 and (4.7) we get for $u\in\mathcal{U}_{ad}$ with $\|y_u-\bar y\|_{C(\bar\Omega)}<\varepsilon$

$$\begin{aligned}
&J'(u)(u-\bar u) = \int_\Omega (\varphi_u+g)(u-\bar u)\,\mathrm{d}x\\
&= \int_\Omega (\varphi_u-\bar\varphi-\psi_{\bar u,u-\bar u})(u-\bar u)\,\mathrm{d}x + \int_\Omega(\bar\varphi+g+\psi_{\bar u,u-\bar u})(u-\bar u)\,\mathrm{d}x\\
&\ge -\frac{\gamma}{2}\|z_{\bar u,u-\bar u}\|^2_{L^2(\Omega)} + [J'(\bar u)(u-\bar u)+J''(\bar u)(u-\bar u)^2] \ge \frac{\gamma}{2}\|z_{\bar u,u-\bar u}\|^2_{L^2(\Omega)}. \qquad \square
\end{aligned}$$

Remark 4.6. *Let us notice that if $\bar u$ is a local minimizer of* (P) *satisfying Assumption 3.4, then there exists $\varepsilon > 0$ such that there is no stationary point $\hat u$ of* (P) *different from $\bar u$ such that $\|y_{\hat u} - \bar y\|_{C(\bar\Omega)} < \varepsilon$. We say that $\hat u$ is a stationary point of* (P) *if it satisfies the first order optimality condition. In particular, if $\hat u$ is a stationary point then $J'(\hat u)(\bar u - \hat u) \geq 0$. This contradicts* (4.6) *if $\|y_{\hat u} - \bar y\|_{C(\bar\Omega)} < \varepsilon$.*

Proof of Theorem 4.4. Using the local optimality of u_ε we get

$$\begin{aligned} 0 &\geq J'_\varepsilon(u_\varepsilon)(u_\varepsilon - \bar u) \\ &= J'(u_\varepsilon)(u_\varepsilon - \bar u) + \int_\Omega \Big[\frac{\partial L}{\partial y}(x, y^\varepsilon_{u_\varepsilon}, u_\varepsilon) - \frac{\partial L}{\partial y}(x, y_{u_\varepsilon}, u_\varepsilon)\Big] z_{u_\varepsilon, u_\varepsilon - \bar u}\,\mathrm{d}x \\ &\quad + \int_\Omega \frac{\partial L}{\partial y}(x, y^\varepsilon_{u_\varepsilon}, u_\varepsilon)(z^\varepsilon_{u_\varepsilon, u_\varepsilon - \bar u} - z_{u_\varepsilon, u_\varepsilon - \bar u})\,\mathrm{d}x + \int_\Omega \eta_\varepsilon z^\varepsilon_{u_\varepsilon, u_\varepsilon - \bar u}\,\mathrm{d}x. \end{aligned} \tag{4.8}$$

We estimate each one of these four terms. First, we observe that the convergence $u_\varepsilon \rightharpoonup \bar u$ in $L^2(\Omega)$ implies that $\|y_{u_\varepsilon} - \bar y\|_{C(\bar\Omega)} \to 0$; see Theorem 2.5. Hence, from Lemma 4.5 we deduce the existence of $\varepsilon_1 > 0$ such that

$$J'(u_\varepsilon)(u_\varepsilon - \bar u) \geq \frac{\gamma}{2}\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|^2_{L^2(\Omega)} \quad \forall \varepsilon < \varepsilon_1, \tag{4.9}$$

For the second term we use Schwarz's inequality, the mean value theorem, (2.8) and (4.2), (3.2), and (4.3)

$$\begin{aligned} &\int_\Omega \Big|\frac{\partial L}{\partial y}(x, y^\varepsilon_{u_\varepsilon}, u_\varepsilon) - \frac{\partial L}{\partial y}(x, y_{u_\varepsilon}, u_\varepsilon)\Big| |z_{u_\varepsilon, u_\varepsilon - \bar u}|\,\mathrm{d}x \\ &\leq C_{L,K_U}\|y^\varepsilon_{u_\varepsilon} - y_{u_\varepsilon}\|_{L^2(\Omega)}\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)} \\ &\leq C_{L,K_U}\sqrt{|\Omega|}C_2\|\xi_\varepsilon\|_{L^2(\Omega)}\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)}. \end{aligned} \tag{4.10}$$

Now we estimate the third term with (3.2) and (4.2), Schwarz's inequality, and (4.4)

$$\begin{aligned} &\int_\Omega \Big|\frac{\partial L}{\partial y}(x, y^\varepsilon_{u_\varepsilon}, u_\varepsilon)\Big| |z^\varepsilon_{u_\varepsilon, u_\varepsilon - \bar u} - z_{u_\varepsilon, u_\varepsilon - \bar u}|\,\mathrm{d}x \leq \int_\Omega \psi_{K_U}|z^\varepsilon_{u_\varepsilon, u_\varepsilon - \bar u} - z_{u_\varepsilon, u_\varepsilon - \bar u}|\,\mathrm{d}x \\ &\|\psi_{K_U}\|_{L^2(\Omega)}C_2^2 C_{f,K_U}\|\xi_\varepsilon\|_{L^2(\Omega)}\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)}. \end{aligned} \tag{4.11}$$

For the last term we use again (4.4) and the fact that $\{\xi_\varepsilon\}_{\varepsilon > 0}$ is bounded in $L^2(\Omega)$

$$\begin{aligned} &\int_\Omega |\eta_\varepsilon z^\varepsilon_{u_\varepsilon, u_\varepsilon - \bar u}|\,\mathrm{d}x \leq \|\eta_\varepsilon\|_{L^2(\Omega)}\Big(\|z^\varepsilon_{u_\varepsilon, u_\varepsilon - \bar u} - z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)} + \|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)}\Big) \\ &\leq \Big(C_2^2\|\xi_\varepsilon\|_{L^2(\Omega)} + 1\Big)\|\eta_\varepsilon\|_{L^2(\Omega)}\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)} \leq C\|\eta_\varepsilon\|_{L^2(\Omega)}\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)}. \end{aligned} \tag{4.12}$$

Inserting the estimates (4.9)–(4.12) in (4.8) we obtain for some constant $C' > 0$ and every $\varepsilon < \varepsilon_1$

$$\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)} \leq C'\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big).$$

Finally, using (2.14) and (4.3) we deduce the the existence of $\varepsilon_2 \in (0, \varepsilon_1]$ such that for every $\varepsilon < \varepsilon_2$ we have

$$\begin{aligned} \|y^\varepsilon_{u_\varepsilon} - \bar y\|_{L^2(\Omega)} &\leq \|y^\varepsilon_{u_\varepsilon} - y_{u_\varepsilon}\|_{L^2(\Omega)} + \|y_{u_\varepsilon} - \bar y\|_{L^2(\Omega)} \\ &\leq C_2\sqrt{|\Omega|}\|\xi_\varepsilon\|_{L^2(\Omega)} + 2\|z_{u_\varepsilon, u_\varepsilon - \bar u}\|_{L^2(\Omega)} \\ &\leq C_2\sqrt{|\Omega|}\|\xi_\varepsilon\|_{L^2(\Omega)} + 2C'\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big), \end{aligned}$$

which proves (4.5). None

5. Stability of the controls. In the previous section, we established Lipschitz stability for the optimal states with respect to state perturbations in the objective functional and to the force in the state equation. In order to obtain stability on the optimal controls an additional assumption is usually required. The reader is referred to [21] for the following assumption

$$\exists C > 0 \text{ such that } |\{x \in \Omega : |(\varphi + g)(x)| \le \varepsilon\}| \le C\varepsilon \quad \forall \varepsilon > 0. \tag{5.1}$$

Using this assumption and sufficient second order optimality conditions they proved Lipschitz stability of the controls in the $L^1(\Omega)$ norm. However, the assumption (5.1) implies that $\bar{u}$ is bang-bang. As far as we know, there is no proof for stability of the optimal controls when they are not bang-bang. Assumption 3.4 that we have considered in the previous sections is applicable for the case of optimal controls that are not bang-bang. Nevertheless, it leads only to Lipschitz stability of the optimal states. Here, we modify Assumption 3.4 as follows

ASSUMPTION 5.1. *There exist numbers $\alpha > 0$ and $\gamma > 0$ such that for all $u \in \mathcal{U}_{ad}$ with $\|y_u - \bar{y}\|_{C(\bar{\Omega})} < \alpha$ the following inequality is fulfilled*

$$J'(\bar{u})(u - \bar{u}) + J''(\bar{u})(u - \bar{u})^2 \ge \gamma \|z_{\bar{u},u-\bar{u}}\|_{L^2(\Omega)} \|u - \bar{u}\|_{L^1(\Omega)}. \tag{5.2}$$

Under this assumption we will prove Lipschitz stability of the optimal controls. We remark that (5.2) does not imply that $\bar{u}$ is bang-bang. Moreover, it has been proved in [9] that the sufficient second order conditions plus the structural assumption (5.1) imply the existence of positive numbers γ and α such that

$$J'(\bar{u})(u-\bar{u}) + J''(\bar{u})(u-\bar{u})^2 \ge \gamma \|u-\bar{u}\|^2_{L^1(\Omega)} \ \forall u \in \mathcal{U}_{ad} \text{ with } \|u-\bar{u}\|_{L^1(\Omega)} < \alpha. \tag{5.3}$$

But we have the next equivalence:

PROPOSITION 5.2. *The statement (5.3) is equivalent to the existence of positive numbers γ' and α' such that*

$$J'(\bar{u})(u-\bar{u}) + J''(\bar{u})(u-\bar{u})^2 \ge \gamma' \|u-\bar{u}\|^2_{L^1(\Omega)} \ \forall u \in \mathcal{U}_{ad} \text{ with } \|y_u-\bar{y}\|_{C(\bar{\Omega})} < \alpha'. \tag{5.4}$$

Proof. Let us assume that (5.3) holds, but (5.4) is false. Then, for every integer $k \ge 1$ there exists an element $u_k \in \mathcal{U}_{ad}$ such that

$$J'(\bar{u})(u_k - \bar{u}) + J''(\bar{u})(u_k - \bar{u})^2 < \frac{1}{k}\|u_k - \bar{u}\|^2_{L^1(\Omega)} \ \text{ and } \ \|y_{u_k} - \bar{y}\|_{C(\bar{\Omega})} < \frac{1}{k}. \tag{5.5}$$

Since $\{u_k\}_{k=1}^\infty \subset \mathcal{U}_{ad}$ is bounded in $L^\infty(\Omega)$, we can extract a subsequence, denoted in the same way, such that $u_k \stackrel{*}{\rightharpoonup} u$ in $L^\infty(\Omega)$. On one side, (5.5) implies that $y_{u_k} \to \bar{y}$ in $C(\bar{\Omega})$. On the other side, from Theorem 2.5 the convergence $y_{u_k} \to y_u$ in $C(\bar{\Omega})$ follows. Then, $y_u = \bar{y}$ and, consequently, $u = \bar{u}$ holds. But (5.3) implies that $\bar{u}$ is bang-bang and, hence, the weak convergence $u_k \stackrel{*}{\rightharpoonup} \bar{u}$ yields the strong convergence $u_k \to \bar{u}$ in $L^1(\Omega)$; see [9, Proposition 12 and Lemma 6]. Then, (5.5) contradicts (5.3).

Let us prove the converse implication. First we observe that given $u \in \mathcal{U}_{ad}$ we get with the mean value theorem

$$\mathcal{A}(y_u - \bar{y}) + \frac{\partial f}{\partial y}(x, \bar{y} + \theta(y_u - \bar{y}))(y_u - \bar{y}) = u - \bar{u}.$$

Now, using (2.2) with $r = 2$ we get

$$\|y_u - \bar y\|_{C(\bar\Omega)} \le C_2\|u - \bar u\|_{L^2(\Omega)} \le C_2\sqrt{u_b - u_a}\|u - \bar u\|^{\frac{1}{2}}_{L^1(\Omega)}.$$

Then, taking $\alpha = \frac{\alpha'^2}{C_2^2(u_b-u_a)}$, we obtain that (5.4) implies (5.3) with $\gamma = \gamma'$. ∎

From (2.3) we infer that (5.4) implies (5.2). Hence, the combination of sufficient second order conditions plus (5.1) is a stronger assumption than (5.2).

THEOREM 5.3. *Let $\bar u$ be a local minimizer of* (P) *satisfying Assumption 5.1 and $\{u_\varepsilon\}_{\varepsilon<\varepsilon_0}$ a family of local solutions of problems* (P_ε) *such that $u_\varepsilon \stackrel{*}{\rightharpoonup} \bar u$ in $L^\infty(\Omega)$ as $\varepsilon \to 0$. Then, there exist $\hat\varepsilon \in (0, \varepsilon_0)$ and a constant $C > 0$ such that*

$$\|u_\varepsilon - \bar u\|_{L^1(\Omega)} \le C\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big) \quad \forall\varepsilon < \hat\varepsilon, \tag{5.6}$$

where $\bar y = y_{\bar u}$.

The proof of this theorem follows the steps of the one of Theorem 4.4 with Lemma 4.5 replaced by the following:

LEMMA 5.4. *Let $\bar u$ satisfy the assumptions of Theorem 5.3. Then, there exists $\varepsilon > 0$ such that*

$$J'(u)(u - \bar u) \ge \frac{\gamma}{2}\|z_{u,u-\bar u}\|_{L^2(\Omega)}\|u - \bar u\|_{L^1(\Omega)}\ \forall u \in \mathcal{U}_{ad} \text{ with } \|y_u - \bar y\|_{C(\bar\Omega)} < \varepsilon, \tag{5.7}$$

where γ is given in Assumption 5.1.

Proof. We use (4.7) with $\rho = \frac{\gamma}{2C_2}$, Assumption 5.1, and (2.3) to deduce for $\varepsilon > 0$ small enough

$$\begin{aligned}
&J'(u)(u - \bar u) = \int_\Omega (\varphi_u + g)(u - \bar u)\,\mathrm{d}x \\
&= \int_\Omega (\varphi_u - \bar\varphi - \psi_{\bar u,u-\bar u})(u - \bar u)\,\mathrm{d}x + \int_\Omega (\bar\varphi + g + \psi_{\bar u,u-\bar u})(u - \bar u)\,\mathrm{d}x \\
&\ge -\frac{\gamma}{2C_2}\|z_{\bar u,u-\bar u}\|^2_{L^2(\Omega)} + [J'(\bar u)(u - \bar u) + J''(\bar u)(u - \bar u)^2] \\
&\ge -\frac{\gamma}{2}\|z_{\bar u,u-\bar u}\|_{L^2(\Omega)}\|u - \bar u\|_{L^1(\Omega)} + \gamma\|z_{\bar u,u-\bar u}\|_{L^2(\Omega)}\|u - \bar u\|_{L^1(\Omega)},
\end{aligned}$$ ∎

which proves (5.7).

Proof of Theorem 5.3. We follow the proof of Theorem 4.4 replacing the estimate (4.9) by (5.7) to deduce with (4.8) and (4.10)–(4.12) the inequality

$$\begin{aligned}
0 \ge J'_\varepsilon(u_\varepsilon)(u_\varepsilon - \bar u) &\ge \frac{\gamma}{2}\|z_{u_\varepsilon,u_\varepsilon-\bar u}\|_{L^2(\Omega)}\|u_\varepsilon - \bar u\|_{L^1(\Omega)} \\
&\quad - C_1\|z_{u_\varepsilon,u_\varepsilon-\bar u}\|_{L^2(\Omega)}\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big).
\end{aligned}$$

Then, dividing this inequality by $\|z_{u_\varepsilon,u_\varepsilon-\bar u}\|_{L^2(\Omega)}$ we get

$$\|u_\varepsilon - \bar u\|_{L^1(\Omega)} \le \frac{2C_1}{\gamma}\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big), \tag{None}$$

which proves (5.6) with $C = \frac{2C_1}{\gamma}$.

6. Some final state stability results. In this section we see how Assumption 5.1 allows us to prove Lipschitz stability for the optimal states for more general perturbations of (P). Here, we consider simultaneous perturbations on the control and state variables of (P):

$$(\mathrm{P}_\varepsilon) \quad \min_{u\in\mathcal{U}_{ad}} J_\epsilon(u) := \int_\Omega L_\varepsilon(x, y_u^\varepsilon(x), u(x))\,\mathrm{d}x,$$

where y_u^ε is the solution of (4.1) and for every $\epsilon > 0$

$$L_\varepsilon(x,y,u) = L_0(x,y) + \eta_\varepsilon y + g_\varepsilon u + \frac{\varepsilon}{2}u^2.$$

As in Section 4, we assume that $\{\xi_\varepsilon\}_{\varepsilon>0}$ and $\{\eta_\varepsilon\}_{\varepsilon>0}$ are bounded families in $L^2(\Omega)$ satisfying that $(\xi_\varepsilon,\eta_\varepsilon)\to(0,0)$ in $L^2(\Omega)^2$ as $\varepsilon\to 0$. Moreover, we suppose that $\|g_\varepsilon - g\|_{L^\infty(\Omega)}\to 0$ as $\varepsilon\to 0$. Under these assumptions, it is immediate to check that (P_ε) is an approximation of (P) in the sense of Theorems 4.2 and 4.3. Moreover, we have the following Lipschitz stability property for the optimal states:

THEOREM 6.1. *Let $\bar u$ be a local minimizer of* (P) *satisfying Assumption 5.1 and $\{u_\varepsilon\}_{\varepsilon<\varepsilon_0}$ a family of local solutions of problems (P_ε) such that $u_\varepsilon \stackrel{*}{\rightharpoonup} \bar u$ in $L^\infty(\Omega)$ as $\varepsilon\to 0$. Then, there exist $\hat\varepsilon\in(0,\varepsilon_0)$ and a constant $C>0$ such that*

$$\|y^\varepsilon_{u_\varepsilon} - \bar y\|_{L^2(\Omega)} \le C\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)} + \|g_\varepsilon - g\|_{L^\infty(\Omega)} + \varepsilon\Big) \quad \forall \varepsilon < \hat\varepsilon, \tag{6.1}$$

where $\bar y = y_{\bar u}$.

Proof. Similarly to (4.8) we have

$$\begin{aligned}0 \ge J'_\varepsilon(u_\varepsilon)(u_\varepsilon-\bar u) &= J'(u_\varepsilon)(u_\varepsilon-\bar u) + \int_\Omega(\varepsilon u_\varepsilon + g_\varepsilon - g)(u_\varepsilon - \bar u)\,\mathrm{d}x\\ &+ \int_\Omega\Big[\frac{\partial L}{\partial y}(x, y^\varepsilon_{u_\varepsilon}, u_\varepsilon) - \frac{\partial L}{\partial y}(x, y_{u_\varepsilon}, u_\varepsilon)\Big] z_{u_\varepsilon, u_\varepsilon-\bar u}\,\mathrm{d}x\\ &+ \int_\Omega \frac{\partial L}{\partial y}(x, y^\varepsilon_{u_\varepsilon}, u_\varepsilon)(z^\varepsilon_{u_\varepsilon,u_\varepsilon-\bar u} - z_{u_\varepsilon,u_\varepsilon-\bar u})\,\mathrm{d}x + \int_\Omega \eta_\varepsilon z^\varepsilon_{u_\varepsilon,u_\varepsilon-\bar u}\,\mathrm{d}x.\end{aligned}$$

Then, using (5.7) and (4.10)–(4.12) we obtain with (2.3)

$$\begin{aligned}0 &\ge \frac{\gamma}{2}\|z_{u_\varepsilon,u_\varepsilon-\bar u}\|_{L^2(\Omega)}\|u_\varepsilon-\bar u\|_{L^1(\Omega)} - \Big(\varepsilon\|u_\varepsilon\|_{L^\infty(\Omega)} + \|g_\varepsilon - g\|_{L^\infty(\Omega)}\Big)\|u_\varepsilon - \bar u\|_{L^1(\Omega)}\\ &- C_1\|z_{u_\varepsilon,u_\varepsilon-\bar u}\|_{L^2(\Omega)}\Big(\|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big) \ge \frac{\gamma}{2}\|z_{u_\varepsilon,u_\varepsilon-\bar u}\|_{L^2(\Omega)}\|u_\varepsilon-\bar u\|_{L^1(\Omega)}\\ &- C'\Big(\varepsilon + \|g_\varepsilon - g\|_{L^\infty(\Omega)} + \|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big)\|u_\varepsilon - \bar u\|_{L^1(\Omega)}.\end{aligned}$$

where $C' = \max\{1, |u_a|, |u_b|, C_1C_2\}$. Dividing the above expression by $\|u_\varepsilon - \bar u\|_{L^1(\Omega)}$ and using (2.14) we infer

$$\|y_{u_\varepsilon} - \bar y\|_{L^2(\Omega)} \le \frac{4C'}{\gamma}\Big(\varepsilon + \|g_\varepsilon - g\|_{L^\infty(\Omega)} + \|\xi_\varepsilon\|_{L^2(\Omega)} + \|\eta_\varepsilon\|_{L^2(\Omega)}\Big).$$ ∎

Now, the rest follows as in the proof of Theorem 4.4

REFERENCES

[1] T. Bayen, F. Bonnans, and F. Silva. Characterization of local quadratic growth for strong minima in the optimal control of semi-linear elliptic equations. *Trans. Amer. Math. Soc.*, 366(4):2063–2087, 2014.

[2] T. Bayen and F. Silva. Second order analysis for strong solutions in the optimal control of parabolic equations. *SIAM J. Control Optim.*, 54(2):819–844, 2016.

[3] E. Casas and M. Mateos. Critical cones for sufficient second order conditions in PDE constrained optimization. *SIAM J. Optim.*, 30(1):585–603, 2020.

[4] E. Casas and M. Mateos. State error estimates for the numerical approximation of sparse distributed control problems in the absence of tikhonov regularization. *Vietnam Journal of Mathematics*, 49:713–738, 2021.

[5] E. Casas, M. Mateos, and A. Rösch. Error estimates for semilinear parabolic control problems in the absence of Tikhonov term. *SIAM J. Control Optim.*, 57(4):2515–2540, 2019.

[6] E. Casas, M. Mateos, and A. Rösch. Analysis of control problems of nonmontone semilinear elliptic equations. *ESAIM Control Optim. Calc. Var.*, 26:Paper No. 80, 21, 2020.

[7] E. Casas, C. Ryll, and F. Tröltzsch. Second order and stability analysis for optimal sparse control of the FitzHugh-Nagumo equation. *SIAM J. Control Optim.*, 53(4):2168–2202, 2015.

[8] E. Casas and F. Tröltzsch. Second order analysis for optimal control problems: Improving results expected from from abstract theory. *SIAM J. Optim.*, 22(1):261–279, 2012.

[9] A. Domínguez Corella, N. Jork, and V. Veliov. Stability in affine optimal control problems constrained by semilinear elliptic partial differential equations. *Submitted*, 2022.

[10] A.L. Dontchev, W.W. Hager, A.B. Poore, and B. Yang. Optimality, stability, and convergence in nonlinear control. *Appl. Math. Optim.*, 31:297–326, 1995.

[11] J.C. Dunn. Second-order optimality conditions in sets of l^∞ functions with range in a polyhedron. *SIAM J. Control Optim.*, 33:1603–1635, 1995.

[12] J.C. Dunn. On second order sufficient conditions for structured nonlinear programs in infinite-dimensional function spaces. In A. Fiacco, editor, *Mathematical Programming with Data Perturbations*, pages 83–107, New York, 1998. Marcel Dekker.

[13] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, Berlin Heidelberg, 1983.

[14] R. Griesse. Lipschitz stability of solutions to some state-constrained elliptic optimal control problems. *Z. Anal. Anwend.*, 25(4):435–455, 2006.

[15] M. Hinze and C. Meyer. Stability of semilinear elliptic optimal control problems with pointwise state constraints. *Comput. Optim. Appl.*, 52(1):87–114, 2012.

[16] B. T. Kien, N. Q. Tuan, C.-F. Wen, and J.-C. Yao. L^∞-stability of a parametric optimal control problem governed by semilinear elliptic equations. *Appl. Math. Optim.*, 84(1):849–876, 2021.

[17] K. Malanowski and F. Tröltzsch. Lipschitz stability of solutions to parametric optimal control for elliptic equations. *Control Cybernet.*, 29(1):237–256, 2000.

[18] H. Maurer. First and second-order sufficient optimality conditions in mathematical programming and optimal control. *Math. Programming Study*, 14:163–177, 1981.

[19] H. Maurer and J. Zowe. First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Math. Programming*, 16:98–110, 1979.

[20] F. Pörner and D. Wachsmuth. Tikhonov regularization of optimal control problems governed by semi-linear partial differential equations. *Math. Control Relat. Fields*, 8(1):315–335, 2018.

[21] N. T. Qui and D. Wachsmuth. Stability for bang-bang control problems of partial differential equations. *Optimization*, 67(12):2157–2177, 2018.

[22] N. S. Trudinger. Linear elliptic operators with measurable coefficients. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (3)*, 27:265–308, 1973.

[23] N. von Daniels. Tikhonov regularization of control-constrained optimal control problems. *Comput. Optim. Appl.*, 70:295–320, 2018.

[24] D. Wachsmuth and G. Wachsmuth. Convergence and regularisation results for optimal control problems with sparsity functional. *ESAIM Control Optim. Calc. Var.*, 17(3):858–886, 2011.

[25] D. Wachsmuth and G. Wachsmuth. Regularization, error estimates and discrepancy principle for optimal control problems with inequality constraints. *Control Cybernet.*, 40(4):1125–1158, 2011.