Metric regularity properties in bang-bang type linear-quadratic optimal control problems

Jakob Preininger, Teresa Scarinci and Vladimir M. Veliov

Metric regularity properties in bang-bang type linear-quadratic optimal control problems[*]

J. Preininger[†], T. Scarinci[‡], and V.M. Veliov[§]

Abstract

The paper investigates the Lipschitz/Hölder stability with respect to perturbations of optimal control problems with linear dynamic and cost functional which is quadratic in the state and linear in the control variable. The optimal control is assumed to be of bang-bang type and the problem to enjoy certain convexity properties. Conditions for bi-metric regularity and (Hölder) metric sub-regularity are established, involving only the order of the zeros of the associated switching function and smoothness of the data. These results provide a basis for the investigation of various approximation methods. They are utilized in this paper for the convergence analysis of a Newton-type method applied to optimal control problems which are affine with respect to the control.

Key words: variational analysis, optimal control, linear control systems, bang-bang controls, metric regularity, stability analysis, Newton's method.
AMS subject classifications: 49J30, 49K40, 49M15, 49N05, 47J07.

1   Introduction

Stability analysis of solutions is a crucial topic in optimization theory due, in particular, to its applications for obtaining error estimates of numerical approximations. Although related investigations in optimal control theory accompany its development from its early stages, the systematic analysis of (Lipschitz) stability in the area started with the works of Dontchev, Hager and Malanowski (see [6, 8]). In these papers, the authors prove Lipschitz dependence of the solutions with respect to perturbations, under a strict coercivity condition which also implies Lipschitz continuity of the optimal control.

In contrast, in the present paper we investigate a class of problems in which the control appears linearly, therefore the strict coercivity fails. Moreover, when the control set is the m-dimensional

[†]Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Austria, `jakob.preininger@tuwien.ac.at`.

[‡]Dept. of Statistics and Operations Research, University of Vienna, Austria, `teresa.scarinci@univie.ac.at`.

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

hypercube $[-1, 1]^m$, each component of the optimal control generally switches from ± 1 to ∓ 1, possibly concatenating with arcs with values in the interior of $[-1, 1]$. That is, the optimal control is typically discontinuous.

Problems which are affine with respect to the control variable arise in many applications, such as engineering, biology and medicine (see e.g. [18, 17, 19, 21]). Nevertheless, only few papers address the stability analysis in case of non-coercive problems and such with discontinuous optimal controls; in fact, many relevant questions still remain unanswered. Recent progress was made in [12, 20, 14, 13] for control-affine problems and in [23] for problems with linear dynamics, and we build on these papers. We mention also the paper [25] and the references therein for problems with group sparsity. Applications to error estimates for time-discretization schemes are discussed in [28, 1, 15, 2, 25, 22] for linear systems or problems of the type (P) below. We mention also the paper [3], where stability analysis is discussed for control-affine systems with bang-singular optimal controls.

In the present paper we focus our attention on the following class of optimal control problems:

$$
\begin{aligned}
\text{minimize} \quad & J(x, u) \\
\text{subject to} \quad & \dot{x}(t) = A(t)x(t) + B(t)u(t) + d(t), \quad t \in [0, T], \\
& u(t) \in U := [-1, 1]^m, \\
& x(0) = x_0,
\end{aligned}
\tag{P}
$$

where

$$
J(x, u) := g(x(T)) + \int_0^T \left(\frac{1}{2} x(t)^\top W(t) x(t) + x(t)^\top S(t) u(t) \right) dt.
$$

Here, $u(t) \in U$ and $x(t) \in \mathbb{R}^n$ denote the control and the state of the system at time $t \in [0, T]$, the function $g : \mathbb{R}^n \to \mathbb{R}$ is given, as well as $A(t), W(t) \in \mathbb{R}^{n \times n}$, $B(t), S(t) \in \mathbb{R}^{n \times m}$ and $d(t) \in \mathbb{R}^n$, $t \in [0, T]$. The set of admissible controls in Problem (P), further denoted by $\mathcal{U}$, consists of all measurable functions u satisfying $u(t) \in U$ for almost every $t \in [0, T]$,

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

Linear terms in u or x are not included in the integrand, which is not a restriction of generality, since such terms can be shifted in a standard way into the differential equation.

The stability properties of the solution(s) of (P) will be analyzed through the Pontryagin minimum principle, which states that for any optimal pair $(\hat{x}, \hat{u})$, there exists an absolutely continuous function $\hat{p} : [0, T] \to \mathbb{R}^n$ such that the triple $(\hat{x}, \hat{p}, \hat{u})$ solves the following system a.e. on $[0, T]$:

$$
\begin{aligned}
0 &= \dot{x}(t) - A(t)x(t) - B(t)u(t) - d(t), \\
0 &= \dot{p}(t) + A(t)^\top p(t) + W(t)x(t) + S(t)u(t), \\
0 &\in B(t)^\top p(t) + S(t)^\top x(t) + N_U(u(t)), \\
0 &= p(T) - \nabla g(x(T)).
\end{aligned}
\tag{PMP}
$$

Here $N_U(u)$ is the normal cone to U at u defined in the usual way:

$$
N_U(u) := \begin{cases} \emptyset & \text{if } u \notin U \\ \{ l \in \mathbb{R}^m : \langle l, v - u \rangle \leq 0 \; \forall v \in U \} & \text{if } u \in U. \end{cases}
$$

2

It will be assumed (see the next sections for precise formulations) that the data are smooth enough, Problem (P) satisfies some convexity-like assumptions, the (reference) optimal control is piece-wise constant with each component taking only the values -1 and 1. Moreover, it will be assumed that each component of the associated "switching function", $t \mapsto B(t)^\top p(t) + S(t)^\top x(t)$, satisfies at its zeros a certain growth condition, characterized by a number $\kappa \geq 1$ (κ can be regarded as the multiplicity of the zeros if the switching function is smooth).

We recast the system (PMP) as the generalized equation

$$0 \in F(x, p, u), \tag{1.1}$$

where F is the set-valued mapping

$$F(x, p, u) := \begin{pmatrix} \dot{x} - Ax - Bu - d \\ \dot{p} + A^\top p + Wx + Su \\ B^\top p + S^\top x + N_{\mathcal{U}}(u) \\ p(T) - \nabla g(x(T)) \end{pmatrix} \tag{1.2}$$

acting in a suitable Banach space $\mathcal{X} \ni (x, p, u)$ with values in a linear normed space $\mathcal{Y}$. The set $N_{\mathcal{U}}(u)$ in (1.2) is a functional replacement for the point-wise cones $N_U(u(t))$ in (PMP) and will be strictly defined in the next section together with the spaces $\mathcal{X}$ and $\mathcal{Y}$.

As usual, we investigate the stability of the solution of problem (P) by introducing a perturbation $y \in \mathcal{Y}$ in the system of necessary optimality conditions, that is, considering the perturbed inclusion $y \in F(x, p, u)$. Under the assumptions briefly mentioned above, the unperturbed system $0 \in F(x, p, u)$, that is the system of necessary optimality conditions (PMP), has a unique solution $(\hat{x}, \hat{p}, \hat{u})$.

Two main concepts of stability are investigated in the paper.

The first concept is a stronger version of the *Hölder strong metric sub-regularity* (see the recent paper [5]). Roughly speaking, we prove that for all sufficiently small perturbations y, the inclusion $y \in F(x, p, u)$, associated with problem (P), has a solution and all the solutions are at distance (in the space $\mathcal{X}$) at most proportional to $\|y\|^{1/\kappa}$ from the unique solution $(\hat{x}, \hat{p}, \hat{u})$ of the inclusion $0 \in F(x, p, u)$. We mention that a similar result was proved in [2, Theorem 9], but with different functional spaces and on slightly stronger assumptions. Moreover, the claim in our result is somewhat stronger, which is rather essential for the analysis of the strong bi-metric regularity and the convergence of Newton's method which will be discussed below.

The second concept extends the standard *strong metric regularity* introduced in the seminal paper [24] by Robinson (see also [9, Chapter 3.7]). The new feature is that a second metric space $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$ is involved (presumably with a non-equivalent and larger metric than that in $\mathcal{Y}$) and only disturbances from this space are considered. Roughly, strong bi-metric regularity relative to $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$ of F at $\hat{z} := (\hat{x}, \hat{p}, \hat{u})$ means that the inverse mapping $\mathcal{Y} \ni y \mapsto F^{-1}(y) = \{z \in \mathcal{X} : y \in F(z)\}$ is locally (around $\hat{z}$) single-valued when restricted to a sufficiently small ball in $\widetilde{\mathcal{Y}}$, centered at $y = 0$. Moreover, this single-valued mapping is Lipschitz continuous with respect to the metric of $\mathcal{Y}$. In the terminology of [9], this means that F has a single-valued localization in $\mathcal{X} \times \widetilde{\mathcal{Y}}$ and it is Lipschitz continuous, but the Lipschitz property holds with respect to the metric of $\mathcal{Y}$.

The general notion of strong bi-metric regularity was introduced in somewhat more restrictive form in [23], where applications to Mayer's type problems for linear control systems were in the focus. Similarly as the strong metric regularity, it has the important property to be invariant with respect to small (in an appropriate sense) functional perturbations of F. This property is often referred to as Lyusternik-Graves type theorem, see e.g. [9, Chapter 5.5]. In the present paper we prove a general Lyusternik-Graves type theorem for strong bi-metrically regular inclusions, which is a substantial improvement of the one in [23], since most of the assumptions are now formulated in terms of the (smaller) metric of $\mathcal{Y}$ rather than in the metric of $\widetilde{\mathcal{Y}}$, as in [23].

We prove strong bi-metric regularity of the mapping F associated with Problem (P), which extends the result in [23] concerning Mayer's problems. This extension is nontrivial, since, technically speaking, the integral cost introduces the state variable in the switching function, making this function nonsmooth. This forces us, among other things, to consider the present slightly more general notion of bi-metric regularity compared with the one in [23]. As an application we give a Lipschitz stability result with respect to small non-linear perturbations in the differential equation.

In the last section of the paper, we investigate the convergence of a Newton-type method (as interpreted in the context of generalized equations, see e.g. [9, Chapter 6.3]) applied to a class of control-affine problems for which (P) can be regarded as a linearization. Notice that the known convergence results (cf. [6]) are inapplicable for non-coercive problems, where the strong metric regularity in the usual space settings fails. We will give sufficient conditions under which the considered Newton's method converges, and does so quadratically. The proof is based on a strengthened version of the metric sub-regularity proved in the present paper for Problem (P). We mention that the stability analysis and the convergence properties of Newton methods still remain not fully understood when singular arcs occur. Some advances have been done recently in [13] for the first issue, and in [3, 11] for the latter. However, these issues remain as interesting topics for future research.

The paper is organized as follows. In Section 2, we recall some basic facts and introduce the main assumptions on Problem (P) together with some notations. Section 3 is devoted to the proof of the Hölder sub-regularity of Problem (P) (actually, of the associated mapping F). In Section 4, we introduce the definition of strong bi-metric regularity, and prove an extension of the Lyusternik-Graves theorem suitable to this new notion. After that, we prove the strong bi-metric regularity of the mapping F resulting from problem (P) and give a result about the invariance of this property under a class of non-linear perturbations. In Section 5, we investigate the convergence of a Newton-type method applied to some control-affine problems with bang-bang solutions.

2 Preliminaries

Throughout the paper we use the following common notations. The standard n-dimensional Euclidean space is denoted by $\mathbb{R}^n$, with the scalar product and norm denoted by $\langle \cdot, \cdot \rangle$ and $|\cdot|$, respectively. The superscript $^\top$ denotes transposition. Further, $L^1([0,T], \mathbb{R}^n)$ and $L^\infty([0,T], \mathbb{R}^n)$ are the spaces of all measurable and absolutely integrable, respectively essentially bounded, functions with the corresponding norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$, which sometimes will be abbreviated as L^1 and L^∞, respectively. Moreover, $W^{1,k}([0,T], \mathbb{R}^n)$ is the space of all absolutely continuous functions from $[0,T]$ to $\mathbb{R}^n$

whose first derivatives belonging to L^k, $k \in \{1, \infty\}$. The corresponding norms are denoted by $\|\cdot\|_{1,1}$ and $\|\cdot\|_{1,\infty}$, respectively. We also denote $W^{1,1}_{x_0}([0,T], \mathbb{R}^n) := \{x \in W^{1,1}([0,T]), \mathbb{R}^n) : x(0) = x_0\}$.

We introduce the following assumptions, some of which will be strengthened in the next sections.

Assumption (A1). The matrix-functions B and S are continuous, A, W and d are measurable and bounded. The matrix $W(t)$ is symmetric for every $t \in [0,T]$. The function g is differentiable with globally Lipschitz continuous gradient ∇g.

We stress that the assumption about *global* Lipschitz continuity of ∇g is made for technical convenience only and is not a real restriction. Since the reachable set in Problem (P) is compact, any modification of g outside a neighborhood of the reachable set does not affect the problem.

For every $u \in \mathcal{U}$ the differential equation in problem (P) with the given initial condition has a unique (absolutely continuous) solution x on $[0,T]$. Every such pair (x,u) is called "admissible", and the set of all admissible pairs is denoted by $\mathcal{F}$.

Thanks to Assumption (A1), a standard compactness argument implies the existence of an optimal solution of Problem (P). In what follows we consider a fixed optimal solution $(\hat{x}, \hat{u})$.

Assumption (A2). For every admissible pair $(x,u) \in \mathcal{F}$ it holds that

$$\langle \nabla g(x(T)) - \nabla g(\hat{x}(T)), \Delta x(T) \rangle + \int_0^T (\langle W(t) \Delta x, \Delta x \rangle + 2 \langle S(t) \Delta u, \Delta x \rangle) dt \geq 0,$$

where $\Delta x(T) := x(T) - \hat{x}(T)$, $\Delta x := x(t) - \hat{x}(t)$ and $\Delta u := u(t) - \hat{u}(t)$.

Let $\hat{p}$ be a co-state function for $(\hat{x}, \hat{u})$, i.e. $(\hat{x}, \hat{p}, \hat{u})$ solves (PMP). We recall that

$$\hat{\sigma} := B^\top \hat{p} + S^\top \hat{x}$$

is the so-called *switching function* corresponding to the triple $(\hat{x}, \hat{p}, \hat{u})$. For every $j \in \{1, \ldots, m\}$ denote by $\hat{\sigma}_j$ its j-th component. Notice that $\hat{\sigma}$ is continuous due to Assumption (A1).

In the next assumption we postulate that the optimal control $\hat{u}$ is *strictly bang-bang*, with a finite number of *switching times* on $[0,T]$, and that the switching function exhibits a certain growth in a neighborhood of any zero.

Assumption (A3). There exist real numbers $\kappa \geq 1$ and $\alpha, \tau > 0$ such that for all $j \in \{1, \ldots, m\}$ and $s \in [0,T]$ with $\hat{\sigma}_j(s) = 0$ we have

$$|\hat{\sigma}_j(t)| \geq \alpha |t - s|^\kappa \quad \forall t \in [s - \tau, s + \tau] \cap [0, T].$$

A similar assumption is introduced in [12] in the case $\kappa = 1$ and in [23, 26] for $\kappa \geq 1$. The set $\mathcal{U}$ of admissible controls will be considered as a metric space with the metric induced by the L^1-norm. For this reason we define

$$\mathcal{X} := W^{1,1}_{x_0}([0,T], \mathbb{R}^n) \times W^{1,1}([0,T], \mathbb{R}^n) \times L^1([0,T], \mathbb{R}^m),$$

with the usual norm: for $(x, p, u) \in \mathcal{X}$,

$$\|(x, p, u)\| := \|x\|_{1,1} + \|p\|_{1,1} + \|u\|_1.$$

Next, we denote by $\mathcal{Y}$ the space

$$\mathcal{Y} := L^1([0,T], \mathbb{R}^n) \times L^1([0,T], \mathbb{R}^n) \times L^\infty([0,T], \mathbb{R}^m) \times \mathbb{R}^n, \tag{2.1}$$

with the usual norm: for $(\xi, \pi, \rho, \nu) \in \mathcal{Y}$,

$$\|(\xi, \pi, \rho, \nu)\| := \|\xi\|_1 + \|\pi\|_1 + \|\rho\|_\infty + |\nu|.$$

We denote by $d_\mathcal{Y}$ the distance induced by $\|\cdot\|$.

As in the introduction, we recast the first order optimality conditions (Pontryagin system) (PMP) for Problem (P) as the generalized equation

$$0 \in F(x, p, u), \tag{2.2}$$

where $F : \mathcal{X} \rightrightarrows \mathcal{Y}$ is defined in (1.2). The normal cone $N_\mathcal{U}(u)$ appearing there is defined in the standard way: for $u \in L^1([0,T], \mathbb{R}^m)$,

$$N_\mathcal{U}(u) := \{v \in L^\infty([0,T], \mathbb{R}^m) : v(t) \in N_U(u(t)) \text{ for a.e. } t \in [0,T]\}.$$

Notice that this definition is consistent with the general definition of a normal cone if $\mathcal{U}$ is considered as a subset of the space L^1 (although $\mathcal{U}$ is also contained in L^∞; but then $N_\mathcal{U}(u)$ should be a cone in the dual space to L^∞).

In the following sections, given a perturbation $y = (\xi, \pi, \rho, \nu) \in \mathcal{Y}$, we will study the inclusion

$$y \in F(x, p, u), \tag{2.3}$$

which, written in detail, looks as follows: for a.e. $t \in [0,T]$,

$$\begin{aligned}
0 &= \dot{x}(t) - A(t)x(t) - B(t)u(t) - d(t) - \xi(t), \\
0 &= \dot{p}(t) + A(t)^\top p(t) + W(t)x(t) + S(t)u(t) - \pi(t), \\
0 &\in B(t)^\top p(t) + S(t)^\top x(t) - \rho(t) + N_U(u(t)), \\
0 &= p(T) - \nabla g(x(T)) - \nu.
\end{aligned} \tag{2.4}$$

3 Strong metric sub-regularity

In this section we prove an important regularity property of the mapping F defined in (1.2), related to, but stronger than, *strong Hölder metric sub-regularity*, see [5].

We begin with some important properties of switching functions that fulfill Assumption (A3). First we fix some notations. Given any continuous function $\sigma : [0,T] \to \mathbb{R}^m$ (σ_j will denote its j-th component) satisfying Assumption (A3) with constants κ, α and τ, and a real number $\delta > 0$, we define

$$I_j(\sigma, \delta) := \bigcup_{s \in [0,T]: \sigma_j(s)=0} (s - \delta, s + \delta) \cap [0,T], \qquad I(\sigma, \delta) := \bigcup_{1 \le j \le m} I_j(\sigma, \delta),$$

and

$$l_{\min}(\sigma, \delta) := \min_{1 \le j \le m} \min_{t \in [0,T] \setminus I_j(\sigma, \delta)} |\sigma_j(t)| > 0. \tag{3.1}$$

Note that this minimum always exists and is indeed positive since σ is continuous and $[0, T] \setminus I_j(\sigma, \tau)$ is compact for any $j \in \{1, \ldots, m\}$.

Now we state an auxiliary result which presents an inverse integral inequality for functions satisfying Assumption (A3) of the type of those developed in Theorem 2.1 and Corollary 2.1 and 2.2 in [27]. It extends [26, Lemma 1.3], which in its turn originates from [12, Lemma 3.3].

Lemma 3.1. *Let $\sigma : [0, T] \to \mathbb{R}^m$ be any continuous function satisfying Assumption (A3). Then there exists a constant $c_0 > 0$ such that*

$$\|v\|_\infty^\kappa \int_0^T \sum_{j=1}^m |\sigma_j(t) v_j(t)| \, dt \geq c_0 \|v\|_1^{\kappa+1} \quad \text{for any } v \in L^\infty([0, T], \mathbb{R}^m). \tag{3.2}$$

Remark 3.2. Carefully following the proof below we can establish that the constant c_0 in the lemma only depends on the numbers κ, α, τ and $l_{\min}(\sigma, \tau)$. Thus Lemma 3.1 can be reformulated in the following more precise form: for any given positive real numbers $\kappa \geq 1$, $\alpha, \tau > 0$ and $m_0 > 0$ there exists a constant $c_0 > 0$ such that the claim (3.2) holds for any continuous function $\sigma : [0, T] \to \mathbb{R}^m$ satisfying Assumption (A3) with constants κ, α, τ, and with $l_{\min}(\sigma, \tau) \geq m_0$.

Proof. If $v = 0$, then the inequality in Lemma 3.1 is fulfilled. If $v \neq 0$ then due to the homogeneity of order $\kappa + 1$ of the two sides of (3.2) with respect to v, it is enough to prove the lemma in the case of $\|v\|_\infty = 1$, which will be assumed in the remaining part of the proof.

Now we choose $\bar{\delta} \in (0, \tau)$ such that $\alpha \bar{\delta}^\kappa < l_{\min}(\sigma, \tau)$. Then for all $\delta \in (0, \bar{\delta}]$ and $j \in \{1, \ldots, m\}$ we have

$$|\sigma_j(t)| \geq \alpha \delta^\kappa \quad \forall t \in [0, T] \setminus I(\sigma, \delta). \tag{3.3}$$

Indeed, if $t \in I_j(\sigma, \tau) \setminus I(\sigma, \delta)$ then inequality (3.3) follows from (A3) and if $t \notin I_j(\sigma, \tau)$ then $|\sigma_j(t)| \geq l_{\min}(\sigma, \tau) > \alpha \bar{\delta}^\kappa \geq \alpha \delta^\kappa$.

Using (3.3) we obtain that

$$\varphi(v) := \int_0^T \sum_{j=1}^m |\sigma_j(t) v_j(t)| \, dt \geq \int_{[0,T] \setminus I(\sigma, \delta)} \sum_{j=1}^m |\sigma_j(t) v_j(t)| \, dt$$

$$\geq \alpha \delta^\kappa \sum_{j=1}^m \int_{[0,T] \setminus I(\sigma, \delta)} |v_j(t)| \, dt \geq \alpha \delta^\kappa \left(\|v\|_1 - \sum_{j=1}^m \int_{I(\sigma, \delta)} |v_j(t)| \, dt \right) \geq \alpha \delta^\kappa (\|v\|_1 - 2\lambda \delta),$$

where λ is the sum of the number of zeros of σ_j for all $j \in \{1, \ldots, m\}$. (Notice that Assumption (A3) implies $\lambda \leq \frac{mT}{2\tau} + m$.) If $\|v\|_1 \geq 4\lambda \bar{\delta}$, we choose $\delta := \bar{\delta}$ to get

$$\varphi(v) \geq \frac{\alpha \bar{\delta}^\kappa}{2} \|v\|_1$$

and since $\|v\|_1 \leq T \|v\|_\infty = T$ we have that $\varphi(v) \geq \frac{\alpha \bar{\delta}^\kappa}{2T^\kappa} \|v\|_1^{\kappa+1}$. If $\|v\|_1 \leq 4\lambda \bar{\delta}$, we choose $\delta := \frac{\|v\|_1}{4\lambda} \leq \bar{\delta}$ to get

$$\varphi(v) \geq \frac{\alpha}{2^{2\kappa+1} \lambda^\kappa} \|v\|_1^{\kappa+1}.$$

Hence, by defining $c_0 := \min\left\{\frac{\alpha \bar{\delta}^\kappa}{2T^\kappa}, \frac{\alpha}{2^{2\kappa+1}\lambda^\kappa}\right\}$ we obtain that

$$\varphi(v) \geq c_0 \|v\|_1^{\kappa+1},$$

which proves (a).

Since we can choose $\bar{\delta}$ to only depend on κ, α, τ and m_0 and there is an upper bound to λ which only depends on m, T and τ, the constant c_0 also only depends on m, T, κ, α, τ and m_0. This proves Remark 3.2. Q.E.D.

The following theorem establishes a stability property of the mapping F associated with system (PMP) which is a somewhat stronger form of the well-known property of *metric sub-regularity*, [9, Section 3I]. It extends [2, Theorem 8] in that Assumption (A3) is weaker than the corresponding assumption there (since we allow 0 and T to be feasible zeros of some components of the switching function), the norm in the space $\mathcal{Y}$ is somewhat weaker, and the function g is not necessarily quadratic. Most importantly, the size of the disturbance y for which the claim of the theorem holds is not a priori restricted (as in the definition of metric sub-regularity, [9, Section 3H] and in [2, Theorem 8]).

Theorem 3.3. *Let $(\hat{x}, \hat{p}, \hat{u})$ be a solution of* (PMP) *such that Assumptions (A1)–(A3) are fulfilled. Then for any $b > 0$ there exists $c > 0$ such that for any $y \in \mathcal{Y}$ with $\|y\| \leq b$, there exists a triple $(x, p, u) \in \mathcal{X}$ solving $y \in F(x, p, u)$, and any such triple satisfies*

$$\|x - \hat{x}\|_{1,1} + \|p - \hat{p}\|_{1,1} + \|u - \hat{u}\|_1 \leq c\|y\|^{\frac{1}{\kappa}}. \tag{3.4}$$

Remark 3.4. Due to further needs, in the proof of the above theorem we will care about how the constant c depends on the data of the problem and the associated switching function $\hat{\sigma}$. More precisely, the following statement will be proved.

Let the natural numbers n, m and the real number $T > 0$ be fixed. Given constants $\kappa \geq 1$, $\alpha > 0$, $\tau > 0$, $m_0 > 0$, $b > 0$ and K, there exists a number $c > 0$ with the following property[1].

Let the $(n \times n)$-matrix functions $A(t)$ and $W(t)$ the $(n \times m)$-matrix functions $B(t)$ and $S(t)$ be defined on $[0,T]$, and $g : \mathbb{R}^n \to \mathbb{R}$ be such that Assumption (A1) is fulfilled, and in addition,

$$\|A\|_\infty, \ \|B\|_\infty, \ \|W\|_\infty, \ \|S\|_\infty, \ \|d\|_\infty, \ \leq K, \quad \nabla g \text{ is Lipschitz with constant } K. \tag{3.5}$$

Let $(\hat{x}, \hat{p}, \hat{u})$ be a solution of (PMP) *(i.e. of (1.1)) such that Assumption (A2) holds, the corresponding switching function $\hat{\sigma}$ fulfills Assumption (A3) with constants κ, α and τ, and $l_{\min}(\hat{\sigma}, \tau) \geq m_0$. Then for every $y \in \mathcal{Y}$ with $\|y\| \leq b$ the inclusion $y \in F(x, p, u)$ (with F defined in (1.2)) has a solution and for every solution (x, p, u) the estimation (3.4) holds.*

Proof. First of all, we note that the inclusion $y \in F(x, p, u)$, for any $y = (\xi, \pi, \rho, \nu) \in \mathcal{Y}$, represents the system of necessary optimality conditions of the following problem:

$$\min\left\{g(x(T)) - \nu^\top x(T) + \int_0^T \left(\frac{1}{2}x(t)^\top W(t)x(t) + x(t)^\top S(t)u(t) - \rho^\top(t)u(t) - \pi^\top(t)x(t)\right)dt\right\} \tag{3.6}$$

---

[1] If $\kappa = 1$, then the constant c can be chosen independent of b.

8

subject to

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + d(t) + \xi(t), \quad t \in [0, T], \quad x(0) = x_0,$$
$$u(t) \in U := [-1, 1]^m.$$

Due to the linearity in u and the convexity and compactness of the constraining set U this problem has a solution, hence also the inclusion $y \in F(x, p, u)$.

Now, let $b > 0$ be arbitrarily chosen and let (x, p, u) be a solution of $y \in F(x, p, u)$, where $y = (\xi, \pi, \rho, \nu) \in \mathcal{Y}$ and $\|y\| \le b$. The following notations will be used. As before, $\hat{\sigma}(t) := B(t)^\top \hat{p}(t) + S(t)^\top \hat{x}(t)$, while $\sigma(t) := B(t)^\top p(t) + S(t)^\top x(t) - \rho(t)$. Furthermore, we denote $\Delta x(t) := x(t) - \hat{x}(t)$, $\Delta p(t) = p(t) - \hat{p}(t)$, $\Delta u(t) := u(t) - \hat{u}(t)$ and $\Delta \sigma(t) := \sigma(t) - \hat{\sigma}(t)$ and skip the argument t whenever this does not lead to ambiguity.

Integrating by parts, we have

$$\int_0^T \langle \Delta \dot{p}, \Delta x \rangle \, dt = \langle \Delta p(T), \Delta x(T) \rangle - \int_0^T \langle \Delta p, \Delta \dot{x} \rangle \, dt.$$

Substituting here the expressions for Δx and Δp resulting from the inclusions $y \in F(x, p, u)$ and $0 \in F(\hat{x}, \hat{p}, \hat{u})$ in view of (1.2) we obtain that

$$\int_0^T \langle -A^\top \Delta p - W \Delta x - S \Delta u + \pi, \Delta x \rangle \, dt$$
$$= \langle \nabla g(x(T)) - \nabla g(\hat{x}(T)) + \nu, \Delta x(T) \rangle - \int_0^T \langle \Delta p, A \Delta x + B \Delta u + \xi \rangle \, dt.$$

Rearranging the terms in this equality and using Assumption (A2) we get

$$\int_0^T (\langle \Delta p, B \Delta u \rangle + \langle S \Delta u, \Delta x \rangle) \, dt + \int_0^T (\langle \pi, \Delta x \rangle + \langle \xi, \Delta p \rangle) \, dt - \langle \nu, \Delta x(T) \rangle$$
$$= \langle \nabla g(x(T)) - \nabla g(\hat{x}(T)), \Delta x(T) \rangle + \int_0^T (\langle W \Delta x, \Delta x \rangle + 2 \langle S \Delta u, \Delta x \rangle) \, dt \ge 0.$$

Using this inequality and the definitions of the functions σ and $\hat{\sigma}$ we obtain

$$\int_0^T \langle \Delta \sigma, \Delta u \rangle \, dt = \int_0^T \langle B^\top \Delta p + S^\top \Delta x - \rho, \Delta u \rangle \, dt \ge$$
$$\ge \int_0^T (-\langle \pi, \Delta x \rangle - \langle \xi, \Delta p \rangle - \langle \rho, \Delta u \rangle) \, dt + \langle \nu, \Delta x(T) \rangle. \qquad (3.7)$$

The third component of the inclusion $y \in F(x, p, u)$ reads as $-\sigma(t) \in N_U(u(t))$, which implies $\langle -\sigma(t), \hat{u}(t) - u(t) \rangle \le 0$. Then

$$-\int_0^T \langle \Delta \sigma, \Delta u \rangle \, dt = \int_0^T [-\langle \sigma, \Delta u \rangle + \langle \hat{\sigma}, \Delta u \rangle] \, dt \ge \int_0^T \langle \hat{\sigma}, \Delta u \rangle \, dt.$$

From here, using that $-\hat{\sigma}_j(t) \in N_{[-1,1]}(\hat{u}_j(t))$, hence $\hat{\sigma}_j(t) \Delta u_j(t) \ge 0$ for each j, Lemma 3.1 implies that

$$-\int_0^T \langle \Delta \sigma, \Delta u \rangle \, dt \ge \int_0^T \sum_{j=1}^m |\hat{\sigma}_j \, \Delta u_j| \, dt \ge c_0 \|\Delta u\|_1^{\kappa+1},$$

9

where the constant c_0 only depends on κ, α, τ and m_0 (see Remark 3.4). Then using (3.7) and the Hölder inequality we obtain

$$\|\pi\|_1 \|\Delta x\|_\infty + \|\xi\|_1 \|\Delta p\|_\infty + |\nu| |\Delta x(T)| + \|\rho\|_\infty \|\Delta u\|_1 \geq c_0 \|\Delta u\|_1^{\kappa+1}. \tag{3.8}$$

Using Assumption (A1) and the solution formula of the Cauchy problem for Δx and Δp we get

$$\|\Delta x\|_\infty \leq c_1(\|\xi\|_1 + \|\Delta u\|_1), \quad \|\Delta p\|_\infty \leq c_2(\|\xi\|_1 + \|\pi\|_1 + \|\Delta u\|_1 + |\nu|) \tag{3.9}$$

for some constants c_1 and c_2 that only depend on K (see (3.5) in Remark 3.4). (We mention that for the estimation of $\|\Delta p\|_\infty$ we use the estimation for $|\Delta x(T)|$ and the Lipschitz continuity of the gradient ∇g appearing in the end-point conditions for p and $\hat{p}$ in (1.1).) Therefore, by (3.8)–(3.9) we obtain that

$$(\|y\|^2 + \|y\|\|\Delta u\|_1) \geq c_3 \|\Delta u\|_1^{\kappa+1} \tag{3.10}$$

for some constant c_3, only depending on c_0, c_1 and c_2. Now, we distinguish two cases. First, if $\|y\| \leq \|\Delta u\|_1$ then

$$2\|y\|\|\Delta u\|_1 \geq c_3 \|\Delta u\|_1^{\kappa+1},$$

which implies

$$\|\Delta u\|_1 \leq \left(\frac{2}{c_3}\|y\|\right)^{1/\kappa}. \tag{3.11}$$

Otherwise, if $\|\Delta u\|_1 \leq \|y\| \leq b$ then

$$\|\Delta u\|_1 \leq \|y\|^{1/\kappa}\|y\|^{(\kappa-1)/\kappa} \leq b^{(\kappa-1)/\kappa}\|y\|^{1/\kappa}. \tag{3.12}$$

Inequalities (3.11) and (3.12) imply that for any $b > 0$ there exists $c_4 > 0$, depending on c_3 and b such that

$$\|\Delta u\|_1 \leq c_4 \|y\|^{1/\kappa}.$$

Then the claim of the theorem follows with a suitable constant c (depending only on c_1, c_2 and c_4) from the above estimation together with (3.9).

Notice that c_4, hence also c, depend on b only due to the term $b^{(\kappa-1)/\kappa}$ in estimation (3.12), which equals 1 in the case $\kappa = 1$. This justifies Footnote 1. $\hfill$ Q.E.D.

Remark 3.5. Clearly, the property established in Theorem 3.3 implies that $(\hat{x}, \hat{p}, \hat{u})$ is the unique solution of (PMP), thus $(\hat{x}, \hat{u})$ is the unique solution of problem (P). Therefore, (PMP), together with Assumptions (A1)–(A3), is a sufficient optimality condition.

4   Bi-metric regularity

The notion of strong bi-metric regularity was introduced in [23] in order to grasp in a relevant way the dependence on perturbations of the solutions of Mayer's type optimal control problems for linear systems. Its extension to the Bolza problem considered in this paper is more complicated due to the missing smoothness of the switching function associated with the optimal control. In this section we present such an extension, starting from the abstract definition of strong bi-metric regularity and a new, substantially strengthened version of the Lyusternik-Graves type theorem proved in [23].

4.1 The abstract setting

First, we give the definition of strong bi-metric regularity, which is a more convenient extension of the one introduced in [23].

Let (X, d_X), (Y, d_Y) and $(\widetilde{Y}, \widetilde{d}_Y)$ be metric spaces, with $\widetilde{Y} \subset Y$ and $d_Y \leq \widetilde{d}_Y$ on $\widetilde{Y}$. Denote by $B_X(\bar{x}; a)$ and $B_{\widetilde{Y}}(\bar{y}; b)$ the closed balls in the metric spaces (X, d_X) and $(\widetilde{Y}, \widetilde{d}_Y)$ with radius $a > 0$ and $b > 0$ centered at $\bar{x}$ and $\bar{y}$, respectively. We will suppose that the metric d_Y and $\tilde{d}_Y$ are *shift-invariant*, which means, in terms of the metric d_Y, that

$$d_Y(y + z, y' + z) = d_Y(y, y'), \quad \forall y, y', z \in Y.$$

Definition 4.1. The map $\Phi : X \rightrightarrows Y$ is strongly bi-metrically regular relative to $\widetilde{Y} \subset Y$ at $\bar{x} \in X$ for $\bar{y} \in \widetilde{Y}$ with constants $\varsigma \geq 0$, $a > 0$ and $b > 0$ if $(\bar{x}, \bar{y}) \in \text{graph}(\Phi)$ and the following properties are fulfilled:

1. the mapping $B_{\widetilde{Y}}(\bar{y}; b) \ni y \mapsto \Phi^{-1}(y) \cap B_X(\bar{x}; a)$ is single-valued, and

2. for all $y, y' \in B_{\widetilde{Y}}(\bar{y}; b)$,

$$d_X(\Phi^{-1}(y) \cap B_X(\bar{x}; a), \Phi^{-1}(y') \cap B_X(\bar{x}; a)) \leq \varsigma d_Y(y, y'). \tag{4.1}$$

It is important to notice that in this definition the "disturbances" y, y' are taken from the smaller space $\widetilde{Y}$ (and are sufficiently small in the metric of this space), but the Lipschitz property (4.1) holds with the (smaller) metric d_Y. This is the crucial difference with the standard definition of strong metric regularity (see e.g. [9, Section 3G] and [16]), where the spaces Y and $\widetilde{Y}$ coincide.

The next result resembles the main features of the Lyusternik-Graves-type theorem proved in [23, Theorem 2.1], but under substantially weakened requirements, as explained in the comments after the proof.

Theorem 4.2. *Let X be a complete metric space, Y be a linear space, $\widetilde{Y}$ be a subspace of Y, and let both metrics, d_Y in Y and $\tilde{d}_Y$ in $\widetilde{Y}$, be shift-invariant and $d_Y \leq \tilde{d}_Y$ on $\widetilde{Y}$. Let the set-valued map $\Phi : X \rightrightarrows Y$ be strongly bi-metrically regular at $\bar{x}$ for $\bar{y}$ with constants ς, a, b. Let $\mu > 0$ and ς' be such that $\varsigma\mu < 1$ and $\varsigma' \geq \varsigma/(1 - \varsigma\mu)$. Then for every positive constants a', b', and γ satisfying*

$$a' \leq a, \quad b' + \gamma \leq b, \quad \varsigma b' \leq (1 - \varsigma\mu)a', \tag{4.2}$$

and for every function $\varphi : X \to \widetilde{Y}$ such that

$$\tilde{d}_Y(\varphi(\bar{x}), \varphi(x)) \leq \gamma \quad \forall x \in B_X(\bar{x}; a'), \tag{4.3}$$

and

$$d_Y(\varphi(x), \varphi(x')) \leq \mu d_X(x, x') \quad \forall x, x' \in B_X(\bar{x}; a'), \tag{4.4}$$

the mapping $B_{\widetilde{Y}}(\bar{y} + \varphi(\bar{x}); b') \ni y \mapsto (\varphi + \Phi)^{-1}(y) \cap B_X(\bar{x}; a')$ is single-valued and Lipschitz continuous with constant ς' with respect to the metric d_Y.

11

Proof. Let us fix μ, ς', a', b' and γ as in the theorem. Take an arbitrary function $\varphi : X \to \widetilde{Y}$ such that (4.3) and (4.4) are fulfilled.

By assumption, the mapping $B_{\widetilde{Y}}(\bar{y}; b) \ni y \mapsto s(y) := \Phi^{-1}(y) \cap B_X(\bar{x}; a)$ is a Lipschitz continuous function (with respect to the metric d_Y in $\widetilde{Y}$) with Lipschitz constant ς. For any $x \in B_X(\bar{x}; a')$ and $y \in B_{\widetilde{Y}}(\bar{y} + \varphi(\bar{x}); b')$ we have

$$\widetilde{d}_Y(y - \varphi(x), \bar{y}) \leq \widetilde{d}_Y(y, \bar{y} + \varphi(\bar{x})) + \widetilde{d}_Y(\varphi(\bar{x}), \varphi(x)) \leq b' + \gamma \leq b. \tag{4.5}$$

Thus $s(y - \varphi(x))$ is defined for all such pairs (x, y).

For an arbitrarily fixed $y \in B_{\widetilde{Y}}(\bar{y} + \varphi(\bar{x}); b')$ we consider the mapping $B_X(\bar{x}; a') \ni x \mapsto Z_y(x) := s(y - \varphi(x))$. We shall prove that the mapping Z_y has a unique fixed point by using the contraction mapping theorem in the form of [9, Theorem 1A.2]. For this we denote $\lambda = \varsigma\mu < 1$ and estimate

$$\begin{aligned} d_X(\bar{x}, Z_y(\bar{x})) &= d_X(s(\bar{y}), s(y - \varphi(\bar{x}))) \leq \varsigma d_Y(\bar{y} + \varphi(\bar{x}), y) \\ &\leq \varsigma b' \leq (1 - \varsigma\mu)a' = (1 - \lambda)a'. \end{aligned}$$

Moreover, for x, $x' \in B_{d_X}(\bar{x}; a')$ we have

$$\begin{aligned} d_X(Z_y(x), Z_y(x')) &= d_X(s(y - \varphi(x)), s(y - \varphi(x'))) \leq \varsigma d_Y(y - \varphi(x), y - \varphi(x')) \\ &= \varsigma d_Y(\varphi(x), \varphi(x')) \leq \varsigma\mu d_X(x, x') = \lambda d_X(x, x'). \end{aligned}$$

Then, according to [9, Theorem 1A.2], there exists a unique $x = x(y) \in B_X(\bar{x}; a')$ such that $x = s(y - \varphi(x))$. The latter implies that $y - \varphi(x) \in \Phi(x)$, hence $x \in (\varphi + \Phi)^{-1}(y) \cap B_X(\bar{x}; a')$. Moreover, $x(y)$ is the unique element of $(\varphi + \Phi)^{-1}(y) \cap B_X(\bar{x}; a')$. Indeed, if $x \in (\varphi + \Phi)^{-1}(y) \cap B_X(\bar{x}; a')$, then $y \in \varphi(x) + \Phi(x)$, hence $y - \varphi(x) \in \Phi(x)$, and since as in (4.5) we have $y - \varphi(x) \in B_{\widetilde{Y}}(\bar{y}; b)$ and $x \in B_X(\bar{x}; a') \subset B_X(\bar{x}; a)$, it also holds that $x = s(y - \varphi(x))$. Thus $x = x(y)$. Thus the mapping $B_{\widetilde{Y}}(\bar{y} + \varphi(\bar{x}); b') \ni y \mapsto (\varphi + \Phi)^{-1}(y) \cap B_X(\bar{x}, a')$ is single-valued.

Now, take two arbitrary elements y, $y' \in B_{\widetilde{Y}}(\bar{y} + \varphi(\bar{x}); b')$ and let $x = s(y - \varphi(x))$ and $x' = s(y' - \varphi(x'))$ be the unique solutions of $y \in \varphi(x) + \Phi(x)$ in $B_X(\bar{x}; a')$ corresponding to y and y', respectively. Then

$$\begin{aligned} d_X(x, x') &= d_X(s(y - \varphi(x)), s(y' - \varphi(x'))) \leq \varsigma d_Y(y - \varphi(x), y' - \varphi(x')) \\ &\leq \varsigma d_Y(y, y') + \varsigma d_Y(\varphi(x), \varphi(x')) \leq \varsigma d_Y(y, y') + \varsigma\mu d_X(x, x'). \end{aligned}$$

Hence,

$$d_X(x, x') \leq \frac{\varsigma}{1 - \varsigma\mu} d_Y(y, y') \leq \varsigma' d_Y(y, y'),$$

which completes the proof. Q.E.D.

The main improvement in the above theorem, compared with [23, Theorem 2.1], is that the Lipschitz property (4.4) is required in [23, Theorem 2.1] to be fulfilled in the stronger metric $\widetilde{d}_Y$, which makes the theorem unusable in several applications, including that presented in Subsection 4.3.

4.2 Strong bi-metric regularity of the linear-quadratic problem

Now consider again Problem (P). First we will present a result about stability under perturbations of Assumption (A3) in the case $\kappa = 1$, where the following strengthened form of Assumption (A1) will be used.

Assumption (A1'). The functions A, W and d are continuous, B and S have continuous first derivatives. The matrices $W(t)$ and $S^\top(t)B(t)$ are symmetric for every $t \in [0,T]$. The function g is differentiable with (globally) Lipschitz continuous gradient.

Furthermore we introduce the subspace

$$\widetilde{\mathcal{Y}} := L^\infty([0,T], \mathbb{R}^n) \times L^\infty([0,T], \mathbb{R}^n) \times W^{1,\infty}([0,T], \mathbb{R}^m) \times \mathbb{R}^n$$

of $\mathcal{Y}$ endowed with the usual norm of $y = (\xi, \pi, \rho, \nu) \in \widetilde{\mathcal{Y}}$:

$$\|(\xi, \pi, \rho, \nu)\|_\sim := \max\{1, T\} \left(\|\xi\|_\infty + \|\pi\|_\infty + \|\rho\|_{1,\infty} + |\nu| \right). \tag{4.6}$$

We denote by $\widetilde{d}_{\mathcal{Y}}$ the distance induced by $\|\cdot\|_\sim$. Notice that $d_{\mathcal{Y}} \leq \widetilde{d}_{\mathcal{Y}}$ on $\widetilde{\mathcal{Y}}$.

Proposition 4.3. *(Stability of Assumption (A3).) Let Assumption (A1') be fulfilled. Let $(\hat{x}, \hat{p}, \hat{u})$ be a solution of* (PMP), *and let Assumption (A2) and Assumption (A3) with $\kappa = 1$ be fulfilled. Then Assumption (A3) is stable under perturbations in the following sense: there exist constants $\widetilde{b} > 0$, $\widetilde{\alpha} > 0$, $\widetilde{\tau} > 0$ and $\widetilde{m}_0 > 0$ such that if $(\xi, \pi, \rho, \nu) = y \in \widetilde{\mathcal{Y}}$ with $\|y\|_\sim \leq \widetilde{b}$, then for any triple $(x, p, u) \in \mathcal{X}$ solving $y \in F(x, p, u)$ the function $\sigma := B^\top p + S^\top x - \rho$ satisfies Assumption (A3) with $\kappa = 1$ and constants $\widetilde{\alpha}$ and $\widetilde{\tau}$ replacing α and τ, respectively, and $l_{\min}(\sigma, \widetilde{\tau}) \geq \widetilde{m}_0$ (see (3.1)).*

Proof. Let α and τ be the constants appearing in Assumption (A3), and let $j \in \{1, \ldots, m\}$ be arbitrary. Further, we consider only disturbances $y \in \widetilde{\mathcal{Y}}$ satisfying $\|y\|_\sim \leq 1$.

First, observe that for all $t \in [0,T]$ it holds that

$$|\sigma_j(t) - \hat{\sigma}_j(t)| \leq \left| \left(B(t)^\top(p(t) - \hat{p}(t)) + S(t)^\top(x(t) - \hat{x}(t)) \right)_j \right| + |\rho_j(t)|.$$

Using this inequality and Theorem 3.3 (applied with $b = 1$), we obtain that there is a constant c_1 such that

$$|\sigma_j(t) - \hat{\sigma}_j(t)| \leq c_1 \|y\|$$

for all $j = 1, \ldots, m$, $t \in [0,T]$, and $y \in \widetilde{\mathcal{Y}}$ with $\|y\|_\sim \leq 1$. Hence,

$$|\sigma_j(t)| \geq |\hat{\sigma}_j(t)| - c_1 \|y\|, \quad t \in [0,T], \quad j \in \{1, \ldots, m\}. \tag{4.7}$$

Consider (skipping the argument t) the derivative

$$\begin{aligned}
\dot{\hat{\sigma}}_j &= \left[\dot{B}^\top \hat{p} + B^\top \dot{\hat{p}} + \dot{S}^\top \hat{x} + S^\top \dot{\hat{x}} \right]_j \\
&= \left[\dot{B}^\top \hat{p} + B^\top(-A^\top \hat{p} - W\hat{x} - S\hat{u}) + \dot{S}^\top \hat{x} + S^\top(A\hat{x} + B\hat{u} + d) \right]_j \\
&= \left[\dot{B}^\top \hat{p} + B^\top(-A^\top \hat{p} - W\hat{x}) + \dot{S}^\top \hat{x} + S^\top(A\hat{x} + d) \right]_j, \tag{4.8}
\end{aligned}$$

13

where in the last inequality we use the symmetricity of $B^\top S$. This implies, in particular, that $\dot{\hat{\sigma}}_j$ is continuous. Then there exists $\tau_1 \in (0, \tau]$ such that $|\dot{\hat{\sigma}}_j(\theta_1) - \dot{\hat{\sigma}}_j(\theta_2)| \leq \alpha/4$ whenever $\theta_1, \theta_2 \in [0, T]$ and $|\theta_1 - \theta_2| < \tau_1$. Hence, using (4.8) and Assumption (A3) we obtain that for any $j \in \{1, \ldots, m\}$, for any zero $\hat{s}$ of $\hat{\sigma}_j$ and arbitrary $t \in (\hat{s} - \tau_1, \hat{s} + \tau_1) \cap [0, T]$

$$\alpha|t - \hat{s}| \leq |\hat{\sigma}_j(t) - \hat{\sigma}_j(\hat{s})| = \left| \int_{\hat{s}}^t \dot{\hat{\sigma}}_j(\theta)\, d\theta \right| \leq \left| \int_{\hat{s}}^t \dot{\hat{\sigma}}_j(\hat{s})\, d\theta \right| + \left| \int_{\hat{s}}^t \frac{\alpha}{4}\, d\theta \right|,$$

hence $|\dot{\hat{\sigma}}_j(\hat{s})| \geq 3\alpha/4$ for any zero $\hat{s}$ of $\hat{\sigma}_j$, $j = 1, \ldots, m$.

The equality (4.8) holds also for σ_j (where $(\hat{x}, \hat{p})$ is replaced with (x, p)), with the additional term $\left[B^\top \pi + S^\top \xi - \dot{\rho} \right]_j$ in the right-hand side. Then using Assumption (A1'), and the estimation in Theorem 3.3, we obtain that

$$\|\dot{\sigma}_j - \dot{\hat{\sigma}}_j\|_\infty \leq c_2(\|y\| + \|\xi\|_\infty + \|\pi\|_\infty + \|\dot{\rho}\|_\infty) \leq c_3\|y\|_\sim, \tag{4.9}$$

where c_2 and c_3 are independent of j and $y \in \widetilde{\mathcal{Y}}$, $\|y\|_\sim \leq 1$.

Define $\widetilde{\tau} := \tau_1/2$ and choose the number $\widetilde{b} > 0$ in such a way that

$$c_1 \widetilde{b} \leq \min\left\{ \frac{l_{\min}(\hat{\sigma}, \widetilde{\tau}/2)}{2}, \frac{\alpha\widetilde{\tau}}{4} \right\} \quad \text{and} \quad 4c_3 \widetilde{b} \leq \alpha, \quad \widetilde{b} \leq 1, \tag{4.10}$$

and let $\|y\|_\sim \leq \widetilde{b}$. Since from (4.7) and the first inequality in (4.10) we have that for $t \in [0, T] \setminus I_j(\hat{\sigma}, \widetilde{\tau}/2)$

$$|\sigma_j(t)| \geq |\hat{\sigma}_j(t)| - c_1\|y\| \geq l_{\min}(\hat{\sigma}, \widetilde{\tau}/2) - c_1\widetilde{b} \geq \frac{l_{\min}(\hat{\sigma}, \widetilde{\tau}/2)}{2} > 0,$$

we obtain that any zero s of σ_j is contained in $I_j(\hat{\sigma}, \widetilde{\tau}/2)$. Thus $s \in (\hat{s} - \widetilde{\tau}/2, \hat{s} + \widetilde{\tau}/2) \cap [0, T]$ for some zero $\hat{s}$ of $\hat{\sigma}_j$.

Now take an arbitrary $t \in (s - \widetilde{\tau}, s + \widetilde{\tau}) \cap [0, T]$. Then $t, s \in (\hat{s} - \tau_1, \hat{s} + \tau_1) \cap [0, T]$ and using (4.9) and the second inequality in (4.10) we obtain that

$$
\begin{aligned}
|\sigma_j(t)| &= \left| \int_s^t \dot{\sigma}_j(\theta)\, d\theta \right| = \left| \int_s^t \left[\dot{\hat{\sigma}}_j(\hat{s}) + (\dot{\hat{\sigma}}_j(\theta) - \dot{\hat{\sigma}}_j(\hat{s})) + (\dot{\sigma}_j(\theta) - \dot{\hat{\sigma}}_j(\theta)) \right] d\theta \right| \\
&\geq |\dot{\hat{\sigma}}_j(\hat{s})||t - s| - \frac{\alpha}{4}|t - s| - c_3\|y\|_\sim |t - s| \\
&\geq \frac{3\alpha}{4}|t - s| - \frac{\alpha}{4}|t - s| - \frac{\alpha}{4}|t - s| \geq \frac{\alpha}{4}|t - s|.
\end{aligned}
$$

Thus (A3) holds for σ with $\kappa = 1$ and constants $\widetilde{\alpha} = \alpha/4$ and $\widetilde{\tau}$.

Further for $t \in I(\hat{\sigma}, \widetilde{\tau}) \setminus I(\sigma, \widetilde{\tau})$ we have

$$|\sigma_j(t)| \geq \alpha|t - \hat{s}| - c_1\|y\| \geq \alpha|t - s| - \alpha|s - \hat{s}| - c_1\|y\| \geq \frac{\alpha\widetilde{\tau}}{4}$$

for some zeros $\hat{s}$ and s of $\hat{\sigma}$ and σ respectively. So if we set $m_0 := \min\{\frac{\alpha\widetilde{\tau}}{4}, l_{\min}(\hat{\sigma}, \widetilde{\tau})\}$ then $l_{\min}(\sigma, \widetilde{\tau}) \geq m_0$. \hfill Q.E.D.

Proposition 4.3 allows to extend the result for strong bi-metric regularity of F, obtained in [23] for Mayer's problems for linear systems, to the present Bolza problem. For that we need the following stronger version of Assumption (A2).

14

Assumption (A2'). For every couple of admissible pairs (x, u), $(x', u') \in \mathcal{F}$ it holds that

$$\langle \nabla g(x(T)) - \nabla g(x'(T)), x(T) - x'(T) \rangle +$$
$$\int_0^T \left(\langle W(t) \left(x(t) - x'(t) \right), x(t) - x'(t) \rangle + 2 \langle S(t) \left(u(t) - u'(t) \right), x(t) - x'(t) \rangle \right) dt \geq 0.$$

Remark 4.4. Standard convex analysis shows that Assumption (A2') is equivalent to the fact that the functional J is convex on the set $\mathcal{F}$, or equivalently, the mapping $L^1([0, T], \mathbb{R}^m) \ni u \mapsto J(x(u), u)$ is convex on the set of admissible controls $\mathcal{U}$, where $x(u)$ denotes the solution of the Cauchy problem $\dot{x} = Ax + Bu$, $x(0) = 0$.

To prove strong bi-metric regularity of (PMP) we first have to introduce the following additional spaces. First we consider the set $\mathcal{U} = L^\infty([0, T], U)$ as a metric space with the metric

$$d^{\#}(u_1, u_2) = \text{meas} \{ t \in [0, T] : u_1(t) \neq u_2(t) \},$$

where "meas" stands for the Lebesgue measure in $[0, T]$. This metric is shift-invariant and we shall shorten $d^{\#}(u_1, u_2) = d^{\#}(u_1 - u_2, 0) =: d^{\#}(u_1 - u_2)$. Moreover, $\mathcal{U}$ is a complete metric space with respect to $d^{\#}$ (see [10, Lemma 7.2]). Then the triple (x, p, u) is considered as an element of the space

$$\widetilde{\mathcal{X}} = W_{x_0}^{1,1}([0, T], \mathbb{R}^n) \times W^{1,1}([0, T], \mathbb{R}^n) \times \mathcal{U}, \tag{4.11}$$

endowed with the (shift-invariant) metric

$$\widetilde{d}_{\mathcal{X}}(x, p, u) = \|x\|_{1,1} + \|p\|_{1,1} + d^{\#}(u). \tag{4.12}$$

Clearly $\widetilde{\mathcal{X}}$ is a complete metric space.

Theorem 4.5 (Bi-metric regularity). *Let Assumptions (A1') and (A2') be fulfilled. Let $(\hat{x}, \hat{p}, \hat{u})$ be a solution of* (PMP) *such that Assumption (A3) is fulfilled with $\kappa = 1$. Then the mapping $F : \widetilde{\mathcal{X}} \rightrightarrows \mathcal{Y}$ introduced in* (1.2) *is strongly bi-metrically regular relative to $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$ at $\hat{z} := (\hat{x}, \hat{p}, \hat{u}) \in \widetilde{\mathcal{X}}$ for $0 \in \widetilde{\mathcal{Y}}$.*

Proof. We shall prove that F^{-1} is single-valued in $B_{\widetilde{\mathcal{Y}}}(0; \widetilde{b})$ and

$$\widetilde{d}_{\mathcal{X}}(F^{-1}(y'), F^{-1}(y)) \leq c d_{\mathcal{Y}}(y', y), \tag{4.13}$$

for all $y, y' \in B_{\widetilde{\mathcal{Y}}}(0; \widetilde{b})$, where $\widetilde{b}$ and c are as in Proposition 4.3. Thus the conditions in Definition 4.1 will be fulfilled even with $a = +\infty$.

Let us start by giving a reformulation of the perturbed version of (P), which will turn out to be useful in the sequel. Let us take an arbitrary $y = (\xi, \pi, \rho, \nu) \in \mathcal{Y}$. Then the perturbed system $y \in F(x, p, u)$ is the set of necessary conditions for the problem (3.6) introduced in the proof of Theorem 3.3. Notice that (3.6) is exactly of the same form as (P) with the state and co-state variables augmented by one dimension, and the data A, B, d, W, S and g replaced with

$$\widetilde{A} = \begin{pmatrix} A & 0 \\ \pi^\top & 0 \end{pmatrix}, \widetilde{B} = \begin{pmatrix} B \\ \rho^\top \end{pmatrix}, \widetilde{d} = \begin{pmatrix} d + \xi \\ 0 \end{pmatrix}, \widetilde{W} = \begin{pmatrix} W & 0 \\ 0 & 0 \end{pmatrix}, \widetilde{S} = \begin{pmatrix} S \\ 0 \end{pmatrix},$$

15

and $\widetilde{g}(x(T), x_{n+1}(T)) = g(x(T)) - \nu^\top x(T) - x_{n+1}(T)$, respectively. Thus, (x, p, u) is a solution of $y \in F(x, p, u)$ if and only if the triple

$$(\widetilde{x}(\cdot), \widetilde{p}(\cdot), \widetilde{u}(\cdot)) = \left(\left(\begin{array}{c} x(\cdot) \\ \int_0^\cdot (\pi^\top x + \rho^\top u) \, dt \end{array} \right), \left(\begin{array}{c} p(\cdot) \\ -1 \end{array} \right), u(\cdot) \right) \tag{4.14}$$

is a solution of the system

$$\begin{aligned}
0 &= \dot{\widetilde{x}}(t) - \widetilde{A}(t)\widetilde{x}(t) - \widetilde{B}(t)\widetilde{u}(t) - \widetilde{d}(t) \\
0 &= \dot{\widetilde{p}}(t) + \widetilde{A}(t)^\top \widetilde{p}(t) + \widetilde{W}(t)\widetilde{x}(t) + \widetilde{S}(t)\widetilde{u}(t) \\
0 &\in \widetilde{B}(t)^\top \widetilde{p}(t) + \widetilde{S}(t)^\top \widetilde{x}(t) + N_U(\widetilde{u}(t)) \\
0 &= \widetilde{p}(T) - \nabla \widetilde{g}(\widetilde{x}(T)).
\end{aligned} \tag{4.15}$$

The above system can be recast as a generalized inclusion

$$0 \in \widetilde{F}_y(\widetilde{x}, \widetilde{p}, \widetilde{u}) \tag{4.16}$$

where $\widetilde{F}_y$ is defined as in (1.2) replacing A by $\widetilde{A}$, and similarly for the other data. $\widetilde{F}_y$ maps the space

$$\hat{\mathcal{X}} := W^{1,1}_{\widetilde{x}_0}([0, T], \mathbb{R}^{n+1}) \times W^{1,1}([0, T], \mathbb{R}^{n+1}) \times L^1([0, T], \mathbb{R}^m)$$

to

$$\hat{\mathcal{Y}} := L^1([0, T], \mathbb{R}^{n+1}) \times L^1([0, T], \mathbb{R}^{n+1}) \times L^\infty([0, T], \mathbb{R}^m) \times \mathbb{R}^{n+1},$$

where $\widetilde{x}_0 := (x_0^\top, 0)^\top$. In few words, the dimension of the state and co-state variable is augmented to $n + 1$ and the additional initial condition $x_{n+1}(0) = 0$ is added.

Note that by construction for any $y \in \widetilde{\mathcal{Y}}$ Assumption (A1) and Assumption (A2') are fulfilled for (4.16). Choose $\widetilde{b}$, $\widetilde{\alpha}$, $\widetilde{\tau}$ and m_0 as in Proposition 4.3. Then there exists a constant K such that for any y with $\|y\|_\sim \leq \widetilde{b}$ we have

$$\|\widetilde{A}\|_\infty, \|\widetilde{B}\|_\infty\|, \|\widetilde{d}\|_\infty, \|\widetilde{W}\|_\infty, \|\widetilde{S}\|_\infty \leq K, \quad \nabla \widetilde{g} \text{ is Lipschitz with constant } K.$$

Then by Proposition 4.3 for any $y = (\xi, \pi, \rho, \nu) \in B_{\widetilde{y}}(0; \widetilde{b})$ and any solution (x, p, u) of the perturbed problem $y \in F(x, p, u)$ Assumption (A3) is satisfied by $\sigma := B^\top p + Sx - \rho$ with constants $\widetilde{\alpha}$, $\widetilde{\tau}$ and $l_{\min}(\sigma, \widetilde{\tau}) \geq m_0$. An easy calculation shows that the switching function of the solution $(\widetilde{x}, \widetilde{p}, \widetilde{u})$ (given by (4.14)) of (4.16) is given by $\widetilde{B}^\top \widetilde{p} + \widetilde{S}^\top \widetilde{x} = B^\top p + S^\top x - \rho = \sigma$. Then Theorem 3.3 in the detailed form in Remark 3.4 is applicable to (4.16) with the constant c independent of the particular $y \in B_{\widetilde{y}}(0; \widetilde{b})$. In particular, this implies that $(\widetilde{x}, \widetilde{p}, \widetilde{u})$ is the unique solution for (4.16). Therefore, $\widetilde{u} = u$ is bang-bang and F^{-1} is single valued on $B_{\widetilde{y}}(0; \widetilde{b})$. For any $y' = (\xi', \pi', \rho', \nu') \in B_{\widetilde{y}}(0; \widetilde{b})$ and its solution (x', p', u') of $y' \in F(x', p', u')$ we define

$$(\widetilde{x}', \widetilde{p}', \widetilde{u}') := ((x', \int_0^\cdot (\pi^\top x' + \rho^\top u')), (p', -1), u'), \quad \widetilde{y}' := ((\xi' - \xi, 0), (\pi' - \pi, 0), \rho' - \rho, (\nu' - \nu, 0)).$$

An easy calculation shows the inclusion $\widetilde{y}' \in \widetilde{F}_y(\widetilde{x}', \widetilde{p}', \widetilde{u}')$. Then Theorem 3.3 (in the form in Remark 3.4) implies

$$\|\widetilde{x}' - \widetilde{x}\|_{1,1} + \|\widetilde{p}' - \widetilde{p}\|_{1,1} + \|\widetilde{u}' - \widetilde{u}\|_1 \leq c\|\widetilde{y}'\|_{\hat{y}}, \tag{4.17}$$

16

where $\| \cdot \|_{\hat{\mathcal{Y}}}$ denotes the norm of $\hat{\mathcal{Y}}$. Hence by (4.17) we have

$$
\begin{aligned}
\|x - x'\|_{1,1} + \|p - p'\|_{1,1} + \|u - u'\|_1 &\leq \|\widetilde{x}' - \widetilde{x}\|_{1,1} + \|\widetilde{p}' - \widetilde{p}\|_{1,1} + \|\widetilde{u}' - \widetilde{u}\|_1 \\
&\leq c\|\widetilde{y}'\|_{\hat{\mathcal{Y}}} = c\|y - y'\|.
\end{aligned}
$$

Since u, u' are bang-bang, similar to [23, p. 4130] we have $\|u - u'\|_1 \geq 2d^{\#}(u - u')$ which proves (4.13).
                                                                                                                        Q.E.D.

We mention that the strong bi-metric regularity for Mayer's problems is proved in [23] for a general polyhedral set U and also in the case $\kappa > 1$. Extension of Theorem 4.5 to a general compact polyhedral U set is a matter of modification of Assumption (A3) and technicalities that we avoid in this paper, while the case $\kappa > 1$ is still open and challenging for the Bolza problem.

4.3 Stability of bi-metric regularity under perturbations

In this subsection, we will apply Theorem 4.2 to prove that the strong bi-metric regularity property is stable under some class of nonlinear perturbations.

Along with problem (P) we consider the following perturbed problem:

$$
\begin{array}{ll}
\text{minimize} & \widetilde{J}(x, u) \\
\text{subject to} & \dot{x}(t) = A(t)x(t) + \widetilde{a}(x(t), t) + B(t)u(t) + \widetilde{B}(x(t), t)u(t), \quad t \in [0, T], \\
& u(t) \in U := [-1, 1]^m, \\
& x(0) = x_0,
\end{array}
\tag{4.18}
$$

where

$$
\begin{aligned}
\widetilde{J}(x, u) := & g(x(T)) + \widetilde{g}(x(T)) + \\
& \int_0^T \left(\frac{1}{2}x(t)^\top W(t)x(t) + \widetilde{w}(x(t), t) + x(t)^\top S(t)u(t) + \langle \widetilde{s}(x(t), t), u(t) \rangle \right) dt.
\end{aligned}
$$

Here $\widetilde{a} : \mathbb{R}^n \times [0, T] \to \mathbb{R}^n$, $\widetilde{B} : \mathbb{R}^n \times [0, T] \to \mathbb{R}^{n \times m}$, $\widetilde{g} : \mathbb{R}^n \to \mathbb{R}$, $\widetilde{w} : \mathbb{R}^n \times [0, T] \to \mathbb{R}$, $\widetilde{s} : \mathbb{R}^n \times [0, T] \to \mathbb{R}^m$ are continuously differentiable functions. All these functions will be assumed "small" in a sense clarified in the theorem below.

The system of necessary optimality conditions for problem (4.18) is given by

$$
\begin{aligned}
0 &= \dot{x}(t) - A(t)x(t) - \widetilde{a}(x(t), t) - B(t)u(t) - \widetilde{B}(x(t), t)u(t), \\
0 &= \dot{p}(t) + \left(A(t) + \widetilde{a}_x(x(t), t) + (\widetilde{B}(x(t), t)u(t))_x\right)^\top p(t) + W(t)x(t) + \widetilde{w}_x(x(t), t)^\top + S(t)u(t), \\
&\quad + \widetilde{s}_x(x(t), t)^\top u(t) \\
0 &\in (B(t) + \widetilde{B}(x(t), t))^\top p(t) + S(t)^\top x(t) + \widetilde{s}(x(t), t) + N_U(u(t)), \\
0 &= p(T) - \nabla g(x(T)) - \nabla \widetilde{g}(x(T)),
\end{aligned}
\tag{4.19}
$$

where the subscript x (as in $\widetilde{a}_x$) means differentiation with respect to x.

The system (4.19) can be recast as

$$
0 \in f(x, p, u) + F(x, p, u),
\tag{4.20}
$$

17

where F (corresponding to the non–perturbed system) is given by (1.2) and f is defined by

$$f(x, p, u)(t) = \begin{pmatrix} -\widetilde{a}(x,t) - \widetilde{B}(x,t)u \\ \left(\widetilde{a}_x(x,t) + (\widetilde{B}(x,t)u)_x\right)^\top p + \widetilde{w}_x(x,t)^\top + \widetilde{s}_x(x,t)^\top u \\ \widetilde{B}(x,t)^\top p + \widetilde{s}(x,t) \\ -\nabla \widetilde{g}(x(T)) \end{pmatrix}.$$

As before we consider F as a set-valued mapping $\widetilde{\mathcal{X}} \rightrightarrows \mathcal{Y}$, where the spaces $\widetilde{\mathcal{X}}$ and $\mathcal{Y}$ are defined in (4.11) and (2.1), respectively. We fix a solution $\hat{z} := (\hat{x}, \hat{p}, \hat{u})$ of the inclusion $0 \in F(x, p, u)$.

Assumption (B). The mapping $F : \widetilde{\mathcal{X}} \rightrightarrows \mathcal{Y}$ is strongly bi-metrically regular relative to $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$ at $\hat{z} \in \widetilde{\mathcal{X}}$ for $0 \in \widetilde{\mathcal{Y}}$.

We recall that sufficient conditions for strong bi-metric regularity of F are given in Theorem 4.5.

Our purpose will be to prove that the strong bi-metric regularity of F is not destroyed by the disturbance f, provided that the disturbances in (4.18) are sufficiently "small". Notice that the space $\widetilde{\mathcal{X}}$ contains elements (x, p, u) for which some of the norms $\|x\|_\infty$, $\|p\|_\infty$, $\|\dot{x}\|_\infty$, $\|\dot{p}\|_\infty$, may be arbitrarily large or even infinite (the latter applies to the derivatives), that is, elements which are irrelevant to the linear-quadratic problem to which F is associated. Moreover, the image $f(\widetilde{\mathcal{X}})$ is not necessarily contained in $\widetilde{\mathcal{Y}}$, which is important from a technical point of view. Therefore, for a given compact set $D \subset \mathbb{R}^n$ we introduce the complete metric space (with the metric $\widetilde{d}_X$)

$$\widetilde{\mathcal{X}}_D := \{(x, p, u) \in \widetilde{\mathcal{X}} : x(t), p(t), \dot{x}(t), \dot{p}(t) \in D \text{ for any } t \in [0, T]\}.$$

Also, denote by $F_D := F|_{\widetilde{\mathcal{X}}_D} : \widetilde{\mathcal{X}}_D \rightrightarrows \mathcal{Y}$ and $f_D := f|_{\widetilde{\mathcal{X}}_D} : \widetilde{\mathcal{X}}_D \to \mathcal{Y}$ the restrictions of F and f to $\mathcal{X}_D$.

Lemma 4.6. *Let Assumption (A1) be fulfilled, let $\hat{z} = (\hat{x}, \hat{p}, \hat{u})$ be a solution of the non-perturbed system* (PMP)*, and let Assumption (B) be fulfilled. Then there exists a compact set $D_0 \subset \mathbb{R}^n$ such that for every compact set $D \subset \mathbb{R}^n$ containing D_0 the restriction f_D maps $\widetilde{\mathcal{X}}_D$ into $\widetilde{\mathcal{Y}}$ and the mapping $F_D : \widetilde{\mathcal{X}}_D \rightrightarrows \mathcal{Y}$ is strongly bi-metrically regular relative to $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$ at $\hat{z} \in \widetilde{\mathcal{X}}$ for $0 \in \widetilde{\mathcal{Y}}$.*

Proof. First note that because of continuity of $\widetilde{a}$, $\widetilde{B}$, $\widetilde{s}$, $\widetilde{a}_x$, $\widetilde{B}_x$, $\widetilde{w}_x$ and $\widetilde{s}_x$ we have that for every compact set D_0 the first three components of f_{D_0} are in L^∞. Moreover the third component is differentiable in t and since $(\widetilde{B}(x,t)^\top p)_x$ is continuous as a function in x, p and t, and $\widetilde{s}_x$ is continuous this derivative lies in L^∞. Hence f_{D_0} maps into $\widetilde{\mathcal{Y}}$.

Further let $\varsigma \geq 0$, $a > 0$ and $b > 0$ be the constants corresponding the strong bi-metric regularity of F. Let $y = (\xi, \pi, \rho, \nu) \in B_{\widetilde{\mathcal{Y}}}(0; b)$ and $(x, p, u) \in \widetilde{\mathcal{X}}$ be a solution the generalized equation $y \in F(x, p, u)$ (i.e. of (2.4)). Moreover we denote $\Delta x(t) := x(t) - \hat{x}(t)$, $\Delta p(t) := p(t) - \hat{p}(t)$ and $\Delta u(t) := u(t) - \hat{u}(t)$. Then by the solution formula of the Cauchy problems for Δx and Δp we get

$$\|\Delta x\|_{1,\infty} \leq c_1(\|\xi\|_\infty + \|\Delta u\|_\infty), \quad \|\Delta p\|_{1,\infty} \leq c_2(\|\xi\|_\infty + \|\pi\|_\infty + \|\Delta u\|_\infty + |\nu|), \qquad (4.21)$$

which shows that there is a compact set D_0 such that $(x, p, u) \in \widetilde{\mathcal{X}}_{D_0}$. Therefore $F^{-1}(B_{\widetilde{\mathcal{Y}}}(0; b)) \subseteq \widetilde{\mathcal{X}}_{D_0} \subseteq \widetilde{\mathcal{X}}_D$ for every D containing D_0 which implies that $F_D : \widetilde{\mathcal{X}}_D \rightrightarrows \mathcal{Y}$ is strongly bi-metrically regular relative to $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$ at $\hat{z} \in \widetilde{\mathcal{X}}$ for $0 \in \widetilde{\mathcal{Y}}$. Q.E.D.

18

Below we prove a stability result in the same spirit as [23, Theorem 4.1], which concerns Mayer's problems. We mention that there is a gap in the proof of [23, Theorem 4.1], but it can be easily corrected by using Theorem 4.2 instead of [23, Theorem 2.1]. This is done in the next theorem which, in addition, extends [23, Theorem 4.1] to Bolza problems.

Theorem 4.7. *Let assumption (A1') be fulfilled, let $\hat{z} = (\hat{x}, \hat{p}, \hat{u})$ be a solution of the non-perturbed system* (PMP), *and let Assumption (B) be fulfilled. Let $D \subset \mathbb{R}^n$ be a compact set such that $f(\widetilde{\mathcal{X}}_D) \subset \widetilde{\mathcal{Y}}$ and the mapping F_D is strongly bi-metrically regular relative to $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$ at $\hat{z} \in \widetilde{\mathcal{X}}_D$ for $0 \in \widetilde{\mathcal{Y}}$ (see Lemma 4.6). Then there exist positive real numbers ε_0, δ and c with the following property.*

For any positive number $\varepsilon \leq \varepsilon_0$ let $\widetilde{a}, \widetilde{B}, \widetilde{g}, \widetilde{w}, \widetilde{s}$ be any functions satisfying the assumptions given above in this section and such that

- *the functions $\widetilde{a}$, $\widetilde{B}$, $\widetilde{s}$, $\widetilde{a}_x$, $\widetilde{B}_x$, $\widetilde{w}_x$, $\widetilde{s}_x$, $\widetilde{B}_t$, $\widetilde{s}_t$ are all bounded by ε on $D \times [0, T]$;*

- *the functions $\widetilde{a}$, $\widetilde{B}$, $\widetilde{s}$, $\widetilde{a}_x$, $\widetilde{B}_x$, $\widetilde{w}_x$, $\widetilde{s}_x$ are Lipschitz continuous in x with Lipschitz constant ε;*

- *the function $\nabla \widetilde{g}$ is bounded by ε and Lipschitz continuous on D with Lipschitz constant ε.*

Then

(i) the perturbed system (4.19) *has a unique solution $z^* = (x^*, p^*, u^*)$ in the δ-neighborhood of $\hat{z}$ in $\widetilde{\mathcal{X}}_D$ and*

$$\widetilde{d}_{\mathcal{X}}(z^* - \hat{z}) \leq c\varepsilon.$$

(ii) the mapping $f + F : \widetilde{\mathcal{X}}_D \rightrightarrows \mathcal{Y}$ is strongly bi-metrically regular at z^* for 0 relative to $\widetilde{\mathcal{Y}} \subset \mathcal{Y}$.

Proof. We want to apply Theorem 4.2 for the mappings $\Phi = F$ and $\varphi = f$ at the point $(\hat{z}, \hat{y})$, where $\hat{y} := f(\hat{x}, \hat{p}, \hat{u})$. Let ς, a, b be the numbers in the definition of strong bi-metric regularity of F at $\hat{z}$ for 0, and let μ, ς', a', b', γ be arbitrary numbers such that the conditions (4.2) are fulfilled.

Since $\widetilde{a}$, $\widetilde{B}$, $\widetilde{s}$, $\widetilde{a}_x$, $\widetilde{B}_x$, $\widetilde{w}_x$, $\widetilde{s}_x$, $\widetilde{B}_t$, $\widetilde{s}_t$, $\nabla \widetilde{g}$ are all bounded by ε and $\dot{\hat{x}}$, $\hat{p}$, $\dot{\hat{p}}$ are bounded by $|D| := \sup_{x \in D} |x|$ and $|\hat{u}| \leq \sqrt{m}$ we have that

$$\widetilde{d}_{\mathcal{Y}}(\hat{y}, 0) = \max\{1, T\} \Big(\| -\widetilde{a}(\hat{x}, t) - \widetilde{B}(\hat{x}, t)\hat{u}\|_\infty + \| \big(\widetilde{a}_x(\hat{x}, t) + (\widetilde{B}(\hat{x}, t)\hat{u})_x\big)^\top \hat{p} + \widetilde{w}_x(\hat{x}, t)^\top + \widetilde{s}_x(\hat{x}, t)^\top \hat{u}\|_\infty$$

$$+ \|\widetilde{B}(\hat{x}, t)^\top \hat{p} + \widetilde{s}(\hat{x}, t)\|_\infty + \| \big(\widetilde{B}(\hat{x}, t)^\top \hat{p}\big)_x \dot{\hat{x}} + \widetilde{B}_t(\hat{x}, t)^\top \hat{p} + \widetilde{B}(\hat{x}, t)^\top \dot{\hat{p}}\|_\infty + |\nabla \widetilde{g}(\hat{x}(T))| \Big)$$

$$\leq \max\{1, T\} \big((1 + \sqrt{m})\varepsilon + (2 + |D| + \sqrt{m})\varepsilon + (|D| + 1)\varepsilon + (|D|^2 + 2|D|)\varepsilon + \varepsilon \big)$$

$$\leq C_1 \varepsilon,$$

for some constant C_1 only depending on $|D|$, m and T. Similarly for $z \in B_{\widetilde{\mathcal{X}}_D}(\hat{z}; a')$ we have

$$\widetilde{d}_{\mathcal{Y}}(0, f(z)) \leq C_1 \varepsilon,$$

which gives

$$\widetilde{d}_{\mathcal{Y}}(\hat{y}, f(z)) \leq \widetilde{d}_{\mathcal{Y}}(\hat{y}, 0) + \widetilde{d}_{\mathcal{Y}}(0, f(z)) \leq 2C_1 \varepsilon. \tag{4.22}$$

Next since $\widetilde{B}$, $\widetilde{a}_x$, $\widetilde{B}_x$, $\widetilde{s}_x$ are bounded by ε and $\widetilde{a}$, $\widetilde{B}$, $\widetilde{s}$, $\widetilde{a}_x$, $\widetilde{B}_x$, $\widetilde{w}_x$, $\widetilde{s}_x$, $\nabla\widetilde{g}$ are Lipschitz continuous with Lipschitz constant ε, p is bounded by $|D|$ and $|u| \leq \sqrt{m}$ we have that for any $z, z' \in B_{\widetilde{\mathcal{X}}_D}(\hat{z}; a')$

$$
\begin{aligned}
d_{\mathcal{Y}}(f(z), f(z')) &= \| -\widetilde{a}(x,t) - \widetilde{B}(x,t)u + \widetilde{a}(x',t) + \widetilde{B}(x',t)u'\|_1 \\
&+ \|\big(\widetilde{a}_x(x,t) + (\widetilde{B}(x,t)u)_x\big)^\top p + +\widetilde{w}_x(x,t)^\top + \widetilde{s}_x(x,t)^\top u \\
&\quad - \big(\widetilde{a}_x(x',t) + (\widetilde{B}(x',t)u')_x\big)^\top p' - \widetilde{w}_x(x',t)^\top - \widetilde{s}_x(x',t)^\top u'\|_1 \\
&+ \|\widetilde{B}(x,t)^\top p + \widetilde{s}(x,t) - \widetilde{B}(x',t)^\top p' - \widetilde{s}(x',t)\|_\infty + |\nabla\widetilde{g}(x(T)) - \nabla\widetilde{g}(x'(T))| \\
&\leq \varepsilon\big[(\|x - x'\|_1 + \sqrt{m}|D|\,\|x - x'\|_1 + \|u - u'\|_1)\|) \\
&+ (|D|\,\|x - x'\|_1 + \|p - p'\|_1 + \sqrt{m}|D|\,\|x - x'\|_1 + |D|\,\|u - u'\|_1 \\
&\quad + \sqrt{m}\|p - p'\|_1 + \|x - x'\|_1 + \sqrt{m}\|x - x'\|_1 + \|u - u'\|_1) \\
&+ (\|x - x'\|_\infty + |D|\,\|x - x'\|_\infty + \|p - p'\|_\infty) + |x(T) - x'(T)|\big] \\
&\leq C_2\varepsilon\|z - z'\|_{\widetilde{\mathcal{X}}_D}
\end{aligned}
$$
(4.23)

for some constant C_2 only depending on $|D|$.

Hence, if we choose ε_0, δ and c such that

$$ 2C_1\varepsilon_0 \leq \gamma,\ C_2\varepsilon_0 \leq \mu,\ C_1\varepsilon_0 < b',\ \delta = a',\ c = \varsigma'C_1,\ c\varepsilon_0 < a', $$

then we can apply Theorem 4.2 to see that $f + F$ is strongly bi-metrically regular at $\hat{z}$ for $\hat{y}$ with constants ς', a' and b'. Therefore, there is a unique $z^* \in B_{\widetilde{\mathcal{X}}_D}(\hat{z}; a')$ such that

$$ 0 \in f(z^*) + F(z^*). $$

and we have

$$ \widetilde{d}_{\mathcal{X}}(z^* - \hat{z}) \leq \varsigma' d_{\mathcal{Y}}(0, \hat{y}) \leq \varsigma' C_1\varepsilon = c\varepsilon, $$

which proves (i). Moreover since $(z^*, 0) \in int(B_{\widetilde{\mathcal{X}}_D}(\hat{z}; a') \times B_{\widetilde{\mathcal{Y}}}(\hat{y}; b'))$, the map $f + F$ is also strongly bi-metrically regular at z^* for 0. This proves (ii).   Q.E.D.

We mention that the issue of stability with respect to linearization of the strong bi-metric regularity property (in the spirit of Robinson's theorem [24]) is more complicated and will be a subject of a separate investigation, together with further applications of this property.

5   A Newton-type method for bang-bang optimal control problems

In this section we investigate the convergence of a Newton-type method for solving affine optimal control problems under conditions which guarantee that the (strengthened) sub-regularity property in Theorem 3.3 holds for the linearized problem along the optimal solution. For this, we first present an abstract result which is similar to, but stronger than [5, Theorem 6.1], since it is based on the stronger version of sub-regularity in Theorem 3.3.

Theorem 5.1. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be Banach spaces. Let the mapping $\varphi : X \to Y$ be Fréchet differentiable ($D\varphi$ denotes the derivative) and let $\Phi : X \rightrightarrows Y$ be a set-valued mapping. Let $\hat{x}$ be a solution of the inclusion*

$$ \varphi(x) + \Phi(x) \ni 0. $$

Assume that there are positive constants R, L and c such that

$$\|D\varphi(x) - D\varphi(\hat{x})\| \leq L\|x - \hat{x}\|_X \quad \forall x \in B_X(\hat{x}, R) \tag{5.1}$$

and

$$\|x - \hat{x}\|_X \leq c\|y\|_Y \tag{5.2}$$

for every $x \in X$ and $y \in \varphi(\hat{x}) + D\varphi(\hat{x})(x - \hat{x}) + \Phi(x)$.

Then for $x \in B_X(\hat{x}, r)$, where $r = \min\{R, \frac{2}{5cL}\}$, and for every solution $z \in X$ of the Newton inclusion

$$\varphi(x) + D\varphi(x)(z - x) + \Phi(z) \ni 0, \tag{5.3}$$

it holds that $z \in B_X(\hat{x}, r)$ and

$$\|z - \hat{x}\|_X \leq \frac{1}{r}\|x - \hat{x}\|_X^2. \tag{5.4}$$

Before proving the theorem we mention that condition (5.2) is a strengthened form of the metric sub-regularity of the partial linearization $x \to \varphi(\hat{x}) + D\varphi(\hat{x})(x - \hat{x}) + \Phi(x)$ of the mapping $\varphi + \Phi$. The inclusion $z \in B_X(\hat{x}, r)$ implies that any finite or infinite sequence generated by the Newton inclusion (5.3) and starting from $B_X(\hat{x}, r)$ (if such exists) stays in $B_X(\hat{x}, r)$. Inequality (5.4) claims quadratic convergence of any such sequence which starts in the interior of $B_X(\hat{x}, r)$.

Proof. For any $x \in B_X(\hat{x}, r)$, let $z \in X$ be an arbitrary solution of (5.3) (if any). Then,

$$\varphi(\hat{x}) + D\varphi(\hat{x})(z - \hat{x}) + \Phi(z) \ni \varphi(\hat{x}) - \varphi(x) + D\varphi(\hat{x})(z - \hat{x}) - D\varphi(x)(z - x).$$

This means that z solves (5.3) with perturbation y given by the right-hand side of the inclusion above. Therefore, (5.2) yields that

$$\|z - \hat{x}\|_X \leq c\|\varphi(\hat{x}) - \varphi(x) + D\varphi(\hat{x})(z - \hat{x}) - D\varphi(x)(z - x)\|_Y.$$

Now using (5.1) we get

$$\begin{aligned}
\|z - \hat{x}\|_X &\leq c(\|\varphi(\hat{x}) - \varphi(x) + D\varphi(\hat{x})(x - \hat{x})\|_Y + \|(D\varphi(\hat{x}) - D\varphi(x))(z - x)\|_Y) \\
&\leq \frac{cL}{2}\|x - \hat{x}\|_X^2 + cL\|x - \hat{x}\|_X \|z - x\|_X \\
&\leq \frac{cL}{2}\|x - \hat{x}\|_X^2 + cL\|x - \hat{x}\|_X (\|z - \hat{x}\|_X + \|x - \hat{x}\|_X).
\end{aligned}$$

Hence,

$$(1 - cL\|x - \hat{x}\|_X)\|z - \hat{x}\|_X \leq \frac{3cL}{2}\|x - \hat{x}\|_X^2.$$

Since $1 - cL\|x - \hat{x}\|_X \geq (1 - cLr) \geq \frac{3}{5}$ we obtain (5.4), which implies that $z \in B_X(\hat{x}, r)$.   Q.E.D.

Remark 5.2. A similar convergence result of the Newton's method can be found in [4] for variational inequalities and nonlinear programming. In that paper, the author introduces the conditions of hemi-stability and hemi-regularity in order to ensure the convergence of the Newton's method. The assumptions in Theorem 5.1 are weaker, but existence of a Newton sequence is not claimed, similarly as to [5, Theorem 6.1]. Existence will follow in the analysis of optimal control problems that follow.

Now, we shall use Theorem 5.1 to investigate the convergence of the Newton method for the following affine optimal control problem:

$$
\begin{aligned}
\text{minimize} \quad & C(x, u) \\
\text{subject to} \quad & \dot{x}(t) = a(x(t), t) + B(x(t), t)u(t), \quad t \in [0, T], \\
& u(t) \in U := [-1, 1]^m, \\
& x(0) = x_0,
\end{aligned}
\tag{5.5}
$$

where

$$
C(x, u) := g(x(T)) + \int_0^T \left[w(x(t), t) + \langle s(x(t), t), u(t) \rangle \right] dt.
$$

Here the functions $a : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$, $B : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^{n \times m}$, $w : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$, $s : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^m$ and $g : \mathbb{R}^n \to \mathbb{R}$ are given. Further, we use the following assumptions.

Assumption (A1"). The functions a, B, w, s are twice differentiable in x, and all these functions and derivatives of first and second order are continuous in t and locally Lipschitz in x, uniformly in t. g is twice continuously differentiable with Lipschitz derivate. The problem (5.5) has a solution, $(\hat{x}, \hat{u})$.

Remark 5.3. The optimality can be understood as local, since it is only important that the Pontryagin maximum principle is fulfilled for $(\hat{x}, \hat{u})$. Due to the linearity of the problem with respect to the control and the compactness and convexity of the control constraints, existence of an optimal solution is granted if the differential equation in (5.5) has a solution on $[0, T]$ for every $u \in \mathcal{U}$.

By the Pontryagin minimum principle, there exists an absolutely continuous function $\hat{p}$ such that the triple $(\hat{x}, \hat{p}, \hat{u})$ solves for a.e. $t \in [0, T]$ the system

$$
\begin{aligned}
& 0 = \dot{x}(t) - a(x(t), t) - B(x(t), t)u(t), \\
& 0 = \dot{p}(t) + \left(a_x(x(t), t) + (B(x(t), t)u(t))_x \right)^\top p(t) + w_x(x(t), t)^\top + s_x(x(t), t)^\top u(t), \\
& 0 \in B(x(t), t)^\top p(t) + s(x(t), t) + N_U(u(t)), \\
& 0 = p(T) - \nabla g(x(T)),
\end{aligned}
\tag{5.6}
$$

where the subscript x (as in a_x) means differentiation with respect to x.

We rewrite system (5.6) as the following generalized equation

$$
0 \in f(x, p, u) + G(x, p, u),
\tag{5.7}
$$

where $f : \mathcal{X} \to \mathcal{Y}$ is given by

$$
f(x, p, u)(t) := \begin{pmatrix} \dot{x} - a(x, t) - B(x, t)u \\ \dot{p} + \left(a_x(x, t) + (B(x, t)u)_x \right)^\top p + w_x(x, t)^\top + s_x(x, t)^\top u \\ B(x, t)^\top p + s(x, t) \\ p(T) - \nabla g(x(T)) \end{pmatrix},
\tag{5.8}
$$

$G : \mathcal{X} \rightrightarrows \mathcal{Y}$ is given by

$$
G(x, p, u) = \begin{pmatrix} 0 \\ 0 \\ N_{\mathcal{U}}(u) \\ 0 \end{pmatrix},
\tag{5.9}
$$

22

and $\mathcal{X}$ and $\mathcal{Y}$ are the spaces defined in Section 2, namely $\mathcal{X} = W_{x_0}^{1,1} \times W^{1,1} \times L^1$, $\mathcal{Y} = L^1 \times L^1 \times L^\infty \times \mathbb{R}^n$.

Following [9, Chapter 6.3], we define the Newton-type method for solving problem (5.7) as follows, where $z^k := (x^k, p^k, u^k)$ denotes the obtained iterate at step $k = 0, 1, \ldots$.

Newton's method:

1. Choose $z^0 \in \mathcal{X}$.

2. Given z^k, obtain z^{k+1} as a solution of the generalized equation

$$f(z^k) + Df(z^k)(z^{k+1} - z^k) + G(z^{k+1}) \ni 0. \tag{5.10}$$

Here, $Df(z)$ is the Jacobian of f at z.

We mention that if z^k satisfies (5.10) then u^k is an admissible control, because $N_{\mathcal{U}}(u) = \emptyset$ whenever $u \notin \mathcal{U}$.

For any $\bar{z} \in \mathcal{X}$ the inclusion $f(\bar{z}) + Df(\bar{z})(z - \bar{z}) + G(z) \ni 0$ represents the Pontryagin system of necessary optimality conditions for a linear-quadratic problem which can be recast as (P) by introducing an additional state variable, similarly in the proof of Theorem 4.5. We denote this problem by $LP(\bar{z})$ (we skip its explicit formulation, which can be found for instance in [7, Section 5]). For the next theorem it is important to ensure that the claim in Theorem 3.3 holds for the particular problem $LP(\hat{z})$ corresponding to $\bar{z} = \hat{z}$, which obviously has the solution $\hat{z}$ – the solution of the non-linearized problem (5.5). Therefore, we make the following assumptions, related to Assumption (A2) and (A3) in Section 2.

Assumption (A2"). The objective functional in problem $LP(\hat{z})$ is convex on the set of all admissible pairs $\mathcal{F}$ (see Remark 4.4).

Assumption (A3"). The switching function $\hat{\sigma}(t)$ in problem $LP(\hat{z})$, which is

$$\hat{\sigma}(t) = B(\hat{x}(t), t)^\top \hat{p}(t) + s(\hat{x}(t), t),$$

satisfies Assumption (A3) with $\kappa = 1$.

The next theorem claims that on the assumptions made, Newton's method generates a sequence quadratically converging to the optimal solution of (5.5).

Theorem 5.4. *Let Assumption (A1") be fulfilled and let $\hat{z} := (\hat{x}, \hat{p}, \hat{u})$ be a solution of problem (5.6). Let, in addition, Assumptions (A2") at (A3") be fulfilled for $\hat{z}$. Then there exists a neighborhood $O \subset \mathcal{X}$ of $\hat{z}$ such that for any starting point $z^0 \in O$ there is a sequence $\{z^k\}_{k=1}^\infty = \{(x^k, p^k, u^k)\}_{k=1}^\infty$ (not necessarily unique) generated by the Newton method (5.10) and any such sequence is quadratically convergent to $\hat{z}$, i.e. there is a constant $c > 0$ such that*

$$\|x^{k+1} - \hat{x}\|_{1,1} + \|p^{k+1} - \hat{p}\|_{1,1} + \|u^{k+1} - \hat{u}\|_1 \leq c \left(\|x^k - \hat{x}\|_{1,1} + \|p^k - \hat{p}\|_{1,1} + \|u^k - \hat{u}\|_1 \right)^2.$$

Proof. Since problem $LP(z^k)$ has a solution and the generalized equation (5.10) represents the Pontryagin necessary optimality conditions for this problem, the iterate z^k exists for every k.

We will apply Theorem 5.1 with spaces $\mathcal{X}$ and $\mathcal{Y}$ (for X and Y) and mappings f and G (for φ and Φ).

An easy but cumbersome calculation (which we skip) shows that Assumption (A1") implies that the mapping $f : \mathcal{X} \to \mathcal{Y}$ is Fréchet differentiable with locally Lipschitz derivative. Thus condition (5.1) in Theorem 5.1 is satisfied with $\varphi = f$ and some constants R and L. Moreover, thanks to Assumptions (A1")–(A3"), Problem $LP(\hat{z})$ fulfills Assumptions (A1)-(A3) in Theorem 3.3. This implies (see Remark 3.4 and Footnote 1) that condition (5.2) in Theorem 5.1 is also fulfilled with some constant c. Then the convergence claimed in the present theorem follows from Theorem 5.1 with the neighborhood O defined as the open ball in $\mathcal{X}$ centered at $\hat{z}$ and with radius r, where r is defined in Theorem 5.1. $\hspace{2cm}$ Q.E.D.

Conclusion

This paper contributes to the regularity theory for Bolza-type optimal control problems with linear dynamics, quadratic in the state and linear in the control objective integrand, and a non-linear terminal term. Conditions for Lipschitz/Hölder sub-regularity and bi-metric regularity are obtained and the results are utilized for obtaining a convergence result for the Newton method applied to non-linear problems that are affine with respect to the control. One of this conditions, which is particularly restrictive, requires that the optimal control is of pure bang-bang type. Extensions of the regularity results and the Newton method to control-affine optimal control problems with singular arcs is an important open area.

References

[1] Alt W., Baier R., Gerdts M., Lempio F.: Approximation of Linear Control Problems with Bang-Bang Solutions. Optimization. 62(1), 9–32 (2013)

[2] Alt W., Schneider C., Seydenschwanz M.: Regularization and implicit Euler discretization of linear-quadratic optimal control problems with bang-bang solutions. Appl. Math. and Comp. 287-288, 104–124 (2016)

[3] Aronna M. S., Bonnans J. F., Martinon P.: A shooting algorithm for optimal control problems with singular arcs. J. Optim. Theory Appl. 158, 419–459 (2013)

[4] Bonnans F. J. : Local analysis of Newton-type methods for variational inequalities and non-linear programming, Appl. Math. Optim., 29, pp. 161–186 (1994)

[5] Cibulka R., Dontchev A.L., Kruger A.Y.: Strong metric subregularity of mappings in variational analysis and optimization. J. Math. Anal. Appl. 457, 1247–1282 (2017)

[6] Dontchev A.L., Hager W.W.: Lipschitzian Stability in Nonlinear Control and Optimization. SIAM J. Control Optim., 31(3), 569–603 (1993)

[7] Dontchev A.L., Hager W.W., Veliov V.M.: Uniform convergence and mesh independence of the Newton method in optimal control. SIAM J. Control and Optim., 39(3), 961–980 (2000)

[8] Dontchev A.L., Malanowski K.: A Characterization of Lipschitzian Stability in Optimal Control. Calculus of variations and optimal control 411, 62–76 (2000)

[9] Dontchev A.L., Rockafellar R.T.: Implicit Functions and Solution Mappings: A View from Variational Analysis. Second edition. Springer, New York (2014)

[10] Ekeland. I.: On the variational principle. J. Math. Anal. and Appl., 47, 324–353 (1974)

[11] Felgenhauer U.: A Newton-type method and optimality test for problems with bang-singular-bang optimal control. Pure Appl. Funct. Anal., 1 (2), 197–215 (2016)

[12] Felgenhauer U.: On Stability of Bang-Bang Type Controls. SIAM J. Control Optim., 41(6), 1843–1867 (2003)

[13] Felgenhauer U.: Stability analysis of variational inequalities for bang-singular-bang controls. Control and Cybernetics, 42(3), 557–592 (2013)

[14] Felgenhauer U., Poggiolini L., Stefani G.: Optimality and stability result for bang-bang optimal controls with simple and double switch behaviour. Control and Cybernetics, 38, 1305–1325 (2009)

[15] Haunschmied J.L., Pietrus A., Veliov V.M.: The Euler Method for Linear Control Systems Revisited. Proceedings of the 9th International Conference on Large-Scale Scientific Computing, 90–97 (2013)

[16] Ioffe A.D.: Metric Regularity. Theory and Applications - a survey. arXiv:1505.07920. `https://arxiv.org/abs/1505.07920`, 24 Oct 2015.

[17] Ledzewicz U., Marriott J., Maurer H., Schättler H.: Realizable protocols for optimal administration of drugs in mathematical models for anti-angiogenic treatment. Mathematical Medicine and Biology: A Journal of the IMA, 27(2), 157–179 (2010)

[18] Ledzewicz U., Schättler H.: Optimal Bang-Bang Controls for a Two-Compartment Model in Cancer Chemotherapy. Journal of Optimization Theory and Applications, 114(3), 609–637 (2002)

[19] Ledzewicz U., Schättler H.: Geometric Optimal Control: Theory, Methods and Examples. Springer, New York (2012)

[20] Maurer H. and Osmolovskii N.P.: Second Order Sufficient Conditions for Time-Optimal Bang-Bang Control SIAM J. Control Optim., 42(6), 2239–2263 (2004)

[21] Maurer H. and Osmolovskii N.P.: Applications to Regular and Bang-Bang Control: Second-Order Necessary and Sufficient Optimality Conditions in Calculus of Variations and Optimal Control. SIAM Advances in Design and Control, vol. 24 (2012)

[22] Pietrus A., Scarinci T., Veliov V.M.: High Order Discrete Approximations to Mayer's Problems for Linear Systems. SIAM J. Control Optim. 56 (1), 102–119 (2018)

[23] Quincampoix M., Veliov V.M.: Metric Regularity and Stability of Optimal Control Problems for Linear Systems. SIAM J. Control Optim., 51(5), 4118–4137 (2013)

[24] Robinson S.M.: Strongly Regular Generalized Equations. Mathematics of Operations Research, 5(1), 43–62 (1980)

[25] Schneider C., Wachsmuth G.: Regularization and discretization error estimates for optimal control of ODEs with group sparsity. ESAIM: Control, Optimisation and Calculus of Variations (2017), doi:10.1051/cocv/2017049

[26] Seydenschwanz M.: Convergence results for the discrete regularization of linear-quadratic control problems with bang-bang solutions. Comput. Optim. Appl., 61(3), 731–760 (2015)

[27] Veliov V.M.: On the Convexity of Integrals of Multivalued Mappings: Applications in Control Theory. Journal of Optimization Theory and Applications, 54(3), 541–563 (1987)

[28] Veliov V.M.: Error analysis of discrete approximations to bang-bang optimal control problems: the linear case. Control and Cybernetics, 34(3), 967–982 (2005)