SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

Karush Kuhn Tucker Conditions for Non-smooth Mathematical Programming Problems in Function Spaces

K. Ito K. Kunisch

SFB-Report No. 2010–038 Dezember 2010

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Karush Kuhn Tucker Conditions for Non-smooth Mathematical Programming Problems in Function Spaces

Kazufumi Ito[1] Karl Kunisch[2]

December 2, 2010

[1]Department of Mathematics, North Carolina State University, Raleigh, North Carolina, 27695-8205, USA; research partially supported by the Army Research Office under DAAD19-02-1-0394

[2]Institut für Mathematik und wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, A-8010 Graz, Austria, supported in part by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 32, "Mathematical Optimization and Applications in the Biomedical Sciences".

Abstract

Lagrange multiplier rules for abstract optimization problems with mixed smooth and convex terms in the cost, with smooth equality constrained and convex inequality constraints are presented. The typical case for the equality constraints that the theory is meant for is given by differential equations. Applications are given to L^1-minimum norm control problems, L^∞-norm minimization, and a class of optimal control problems with distributed state constraints and non-smooth cost.

1 Introduction

In this paper we discuss non-smooth mathematical programming problems, which in part rely on convexity, to the other part on regularity assumptions, more precisely we consider

$$(P) \qquad \begin{cases} \min\ F(y) = F_0(y) + F_1(y) \\ \\ \text{subject to } G_1(y) = 0, \;\; G_2(y) \le 0,\, y \in \mathcal{C}, \end{cases}$$

where F_0 and G_1 are C^1 mappings, F_1 and G_2 are convex, and $\mathcal{C}$ denotes a set of additional constraints. Since the range spaces of G_1, G_2 are not necessarily finite dimensional, the notion of convexity will have to be made clear. The focus on the research lies on the derivation of optimality conditions which can be expressed as equations rather then differential inclusions. This can be achieved by means of Lagrange multipliers. One motivation for this procedure is given by the fact that that nonlinear equations are simpler to realize numerically than differential inclusions. If all operations in (P) are smooth, then the Maurer-Zowe-Kurcyusz [MaZo], [ZoKu] conditions provide the Lagrangian framework that we are looking for. In the case that non-differentiable terms arise in the problem formulation, an analogously general framework does not appear to be available. The typical case for the equality constraints that we have in mind is given by differential equations. Typical cases for F_1 are L^1 and L^∞ type functionals. The former arise in the context of sparse controls, see e.g. [CK, WW] and L^1 data-fitting, see e.g. [CJK] and the references given there, the latter, for example in the context of minimal effort optimal control problems.

The approach that we follow here to derive Lagrange multiplier rules essentially rests on the use of Ekeland's variational principle [E]. It has been used in a series of papers focusing on state-constrained optimal control problems, [C, CY, LY]. As we shall demonstrate, however, the technique is general, and quite constructive, in the sense that the Lagrange multipliers are the limits of expression resulting from asymptotic expressions resulting from the variational principle. The approach follows the line of the possibly non-qualified form, i.e. there appears also a multiplier associated to the cost F. Special attention must made to guarantee that at least one of the multipliers is nontrivial, and in particular, that the multiplier associated to F is nontrivial.

We next briefly outline the contents of the paper. In Section 2 the case were the range spaces of G_1 and G_2 are finite-dimensional is treated. The case of infinite dimensional image spaces is considered in Section 3. To guarantee non-triviality of the multipliers, we need regularity conditions are. The conditions that we utilize can be seen as generalizations of the Maurer-Zowe-Kurcyusz conditions from the smooth to the convex case. The following three sections are devoted to three applications: L^1-minimum norm control problems, which arise in the context of optimal control with sparsity constraints, L^∞-norm minimization, and a class of optimal control problems with distributed state constraints and nonsmooth cost. The applications presented here do not aim for strongest generality and can certainly be extended in future work.

2 Finite Dimensional Range Case

In this section we investigate (P) for the case where the range spaces of the mappings G_1 and G_2 are finite dimensional. The following assumption will be supposed to hold troughout.

$$\text{(H1)} \quad \begin{cases} F = F_0(y) + F_1(y), \\ \text{where } F_0 \in C^1(X, \mathbb{R}) \text{ and } F_1 : X \to \mathbb{R} \cup \{\infty\} \\ \text{is convex and continuous on the effective domain,} \\ G_1 \in C^1(X, \mathbb{R}^m), \\ G_2 : X \in C(X, \mathbb{R}^p), \text{ with } (G_2)_i \text{ convex }, \\ \mathcal{C} \subset X \text{ is closed and convex.} \end{cases}$$

Here X denotes a real Banach space. The effective domain $\{v \in X : F_1(v) < \infty\}$ will be denoted by $domF_1$. We have the following necessary optimality condition.

Theorem 2.1. *Let* (H1) *hold and let* $y^* \in \mathcal{C}$ *be a local minimum of* (P).

Then there exists a nontrivial $(\lambda_0, \mu_1, \mu_2) \in \mathbb{R}^+ \times \mathbb{R}^m \times \mathbb{R}^p$ *such that*
(2.1)

$$\begin{cases} \lambda_0 \left(F_0'(y^*)(y-y^*) + F_1(y) - F_1(y^*)\right) + (\mu_1, G_1'(y^*)(y-y^*))_{\mathbb{R}^m} \\ \qquad\qquad +(\mu_2, G_2(y) - G_2(y^*))_{\mathbb{R}^p} \geq 0 \textit{ for all } y \in \mathcal{C} \cap domF_1, \\ \mu_2 \geq 0, \quad G_2(y^*) \leq 0, \quad (\mu_2, G_2(y^*))_{\mathbb{R}^p} = 0. \end{cases}$$

Proof. For $\epsilon > 0$ define the regularized functional

$$J_\epsilon(y) = \left(((F(y) - F(y^*) + \epsilon)^+)^2 + |G_1(y)|^2_{\mathbb{R}^m} + |\max(0, G_2(y))|^2_{\mathbb{R}^p}\right)^{\frac{1}{2}}.$$

Then J_ϵ is continuous on $domF_1$ and $\epsilon = J_\epsilon(y^*) \leq \inf\ J_\epsilon + \epsilon$. The norm on X in a natural way defines a metric on $\mathcal{C}$ by means of $d(y_1, y_2) = |y_1 - y_2|_X$. By the Ekeland variational principle, see e.g. [Cl], pg. 266, there exists a $y^\epsilon \in \mathcal{C}$ such that

$$\begin{cases} J_\epsilon(y^\epsilon) \leq J_\epsilon(y^*), \\ J_\epsilon(y) - J_\epsilon(y^\epsilon) \geq -\sqrt{\epsilon}\, d(y, y^\epsilon) \quad \text{for all } y \in \mathcal{C}, \\ d(y^\epsilon, y^*) \leq \sqrt{\epsilon}. \end{cases} \tag{2.2}$$

Choose $\hat{y} \in \mathcal{C} \cap domF_1$. Throughout we assume that $\epsilon \in (0, 1)$. Then in particular $\{y^\epsilon : \epsilon \in (0, 1)\}$ is bounded. We further set

$$y_t = y^\epsilon + t\,(\hat{y} - y^\epsilon),\ t \in (0, 1),$$

and

$$\tilde{\mu}_1^\epsilon = G_1(y^\epsilon) \in \mathbb{R}^m, \quad \tilde{\mu}_2^{\epsilon,t} = \max(0, G_2(y_t)) \in \mathbb{R}^p. \tag{2.3}$$

By convexity of $\mathcal{C}$ and F_1 it follows that y_t is in the effective domain of F_1 and $y_t \in \mathcal{C}$. Setting $y = y_t$ in (2.2), we have

$$-\sqrt{\epsilon}\, d(y_t, y^\epsilon) \leq J_\epsilon(y_t) - J_\epsilon(y^\epsilon). \tag{2.4}$$

For the following estimate we use

$$\sqrt{\sum |a_i|^2} - \sqrt{\sum |b_i|^2} = \frac{1}{\sqrt{\sum |a_i|^2} + \sqrt{\sum |b_i|^2}} \sum (a_i + b_i, a_i - b_i).$$

Then from (2.4)

(2.5)
$$
\begin{aligned}
&-\sqrt{\epsilon}\, d(y_t, y^\epsilon) \\
&\le \tfrac{1}{J_\epsilon(y^\epsilon)+J_\epsilon(y_t)} \big[\tilde\alpha^{\epsilon,t} \big((F(y_t) - F(y^*) + \epsilon)^+ - (F(y^\epsilon) - F(y^*) + \epsilon)^+\big) \\
&\qquad + \big(|G_1(y_t)|^2_{\mathbb{R}^m} - |G_1(y^\epsilon)|^2_{\mathbb{R}^m}\big) + \big(|\max(0, G_2(y_t))|^2_{\mathbb{R}^p} - |\max(0, G_2(y^\epsilon))|^2_{\mathbb{R}^p}\big)\,\big],
\end{aligned}
$$

where
$$\tilde\alpha^{\epsilon,t} = ((F(y_t) - F(y^*) + \epsilon)^+ + (F(y^\epsilon) - F(y^*) + \epsilon)^+).$$

The three additive terms on the right hand side of (2.5) are considered next. For G_1 we use that for every $\eta > 0$ there exists $\delta > 0$ such that

$$\|G_1'(y) - G_1'(y^*)\| < \frac{\eta}{2} \;\text{ if }\; |y - y^*| < \delta.$$

As a consequence there exist $\epsilon(\eta)$ and $t(\eta)$ such that

$$\|G_1'(y_t) - G_1'(y^*)\| < \eta \;\text{ for all } \epsilon \in (0, \epsilon(\eta)),\; t \in (0, t(\eta)). \tag{2.6}$$

We can choose $\epsilon(\eta)$ and $t(\eta)$ such that in addition

$$\|F_0'(y_t) - F_0'(y^*)\| < \eta \;\text{ for all } \epsilon \in (0, \epsilon(\eta)),\; t \in (0, t(\eta)). \tag{2.7}$$

For every $\epsilon \in (0,1)$ there exists $\bar t(\epsilon) > 0$ such that for all $t \in (0, \bar t(\epsilon))$

$$(F(y_t) - F(y^*) + \epsilon)(F(y^\epsilon) - F(y^*) + \epsilon) \ge 0.$$

Together with convexity of F_1 this implies that for $t \in (0, \bar t(\epsilon))$

(2.8)
$$
\begin{aligned}
&\tilde\alpha^{\epsilon,t} \big((F(y_t) - F(y^*) + \epsilon)^+ - (F(y^\epsilon) - F(y^*) + \epsilon)^+\big) \le \tilde\alpha^{\epsilon,t}(F(y_t) - F(y^\epsilon)) \\
&\quad \le \tilde\alpha^{\epsilon,t}\big(t \textstyle\int_0^1 F_0'(s\, y_t + (1-s)y^\epsilon)(\hat y - y^\epsilon)\, ds + t\, F_1(\hat y - y^\epsilon)\big) \\
&\quad \le \tilde\alpha^{\epsilon,t}\big(\, tF_0'(y^\epsilon)(\hat y - y^\epsilon) + t\, F_1(y^\epsilon)(\hat y - y^\epsilon) + o(|y_t - y^\epsilon|)\big)
\end{aligned}
$$

where in the last estimate we used (2.7), and

$$t(\hat y - y^\epsilon) = y_t - y^\epsilon \text{ and } sy_t + (1-s)y^\epsilon = y_{st}.$$

For the second term on the right hand side of (2.5) we find

$$
\begin{aligned}
&|G_1(y_t)|^2_{\mathbb{R}^m} - |G_1(y^\epsilon)|^2_{\mathbb{R}^m} = \big(G_1(y_t) + G_1(y^\epsilon), G_1(y_t) - G_1(y^\epsilon)\big)_{\mathbb{R}^m} \\
&\quad = \big(2G_1(y^\epsilon), G_1(y_t) - G_1(y^\epsilon)\big)_{\mathbb{R}^m} + |G_1(y_t) - G_1(y^\epsilon)|^2_{\mathbb{R}^m} \le \big(2G_1(y^\epsilon), G_1'(y^\epsilon)(y_t - y^\epsilon)\big)_{\mathbb{R}^m} \\
&\quad +2\|G_1(y^\epsilon)\|_{\mathbb{R}^m} \int_0^1 \|G_1'(s\,y_t + (1-s)y^\epsilon) - G_1'(y^\epsilon)\|_{\mathbb{R}^m}\, ds\, |y_t - y^\epsilon|_X + |G_1(y_t) - G_1(y^\epsilon)|^2_{\mathbb{R}^m} \\
&\quad \le 2(\tilde{\mu}_1^\epsilon, G_1'(y^\epsilon)(y_t - y^\epsilon))_{\mathbb{R}^m} + o(|y_t - y^\epsilon|_X) = 2(\tilde{\mu}_1^\epsilon, G_1'(y^\epsilon)(y_t - y^\epsilon))_{\mathbb{R}^m} + o(|t|),
\end{aligned}
$$

where in the last inequality we used (2.6). Turning to the third term on the right hand side of (2.5) we estimate using convexity of $v \to |\max(0,v)|^2_{\mathbb{R}^p}$

$$
\begin{aligned}
&|\max(0, G_2(y_t))|^2_{\mathbb{R}^p} - |\max(0, G_2(y^\epsilon))|^2_{\mathbb{R}^p} \\
&\quad \le \big(2\max(0, G_2(y_t)), G_2(y_t) - G_2(y^\epsilon)\,\big)_{\mathbb{R}^p} \le t\,\big(\mu_2^{\epsilon,t}, G_2(\hat{y}) - G_2(y^\epsilon)\,\big)_{\mathbb{R}^p},
\end{aligned}
$$

where in the last estimate we used the coordinate-wise convexity of G_2 and the notation introduced in (2.3). Combining these estimates we arrive at

(2.9)

$$
\begin{aligned}
&-\sqrt{\epsilon}\, d(y_t, y^\epsilon) \\
&\quad \le \tfrac{1}{J_\epsilon(y^\epsilon)+J_\epsilon(y_t)} \big[\tilde{\alpha}^{\epsilon,t}\big(tF_0'(y^\epsilon)(\hat{y} - y^\epsilon) + t(F_1(\hat{y}) - F_1(y^\epsilon))\big) + \\
&\qquad 2\,(\tilde{\mu}_1^\epsilon, G_1'(y^\epsilon)(y_t - y^\epsilon))_{\mathbb{R}^m} + 2\,t\,(\tilde{\mu}_2^{\epsilon,t}, G_2(\hat{y}) - G_2(y^\epsilon)\,)_{\mathbb{R}^p}\big] + o\,(|t|),
\end{aligned}
$$

for all $t \in (0, \bar{t}(\epsilon))$. Let

$$
\alpha^{\epsilon,t} = \frac{\tilde{\alpha}^{\epsilon,t}}{J_\epsilon(y_t) + J_\epsilon(y^\epsilon)}, \quad \mu^{\epsilon,t} = \frac{2\,\tilde{\mu}^{\epsilon,t}}{J_\epsilon(y_t) + J_\epsilon(y^\epsilon)}, \quad \mu^{\epsilon,t} = (\mu_1^\epsilon, \mu_2^{\epsilon,t}), \quad \mu_2^{\epsilon,t} \ge 0.
$$

Taking the limit as $t \to 0^+$ implies that

$$
\lambda_{0,\epsilon} := \lim_{t\to 0^+} \alpha^{\epsilon,t} = \tfrac{(F(y^\epsilon) - F(y^*) + \epsilon)^+}{J_\epsilon(y^\epsilon)} \text{ in } \mathbb{R},
$$

and

$$
\mu^\epsilon := \lim_{t\to 0^+} \mu^{\epsilon,t} = \tfrac{\big(G_1(y^\epsilon),\, \max(0, G_2(y^\epsilon))\big)}{J_\epsilon(y^\epsilon)} \text{ in } \mathbb{R}^m \times \mathbb{R}^p.
$$

We have $\lambda_{0,\epsilon} \ge 0$, $\mu_2^\epsilon \ge 0$ and

$$
|(\lambda_{0,\epsilon}, \mu^\epsilon)|_{\mathbb{R}\times\mathbb{R}^m\times\mathbb{R}^p} = 1. \tag{2.10}
$$

Dividing (2.9) by t and letting $t \to 0^+$, we obtain

$$-\sqrt{\epsilon} d(\hat{y}, y^\epsilon) \leq \lambda_{0,\epsilon} \left(F_0'(y^\epsilon)(\hat{y} - y^\epsilon) + F_1(\hat{y}) - F_1(y^\epsilon)\right)$$

$$+(\mu_1^\epsilon, G_1'(y^\epsilon)(\hat{y} - y^\epsilon))_{\mathbb{R}^m} + (\mu_2^\epsilon, G_2(\hat{y}) - G_2(y^\epsilon))_{\mathbb{R}^p}.$$

Since $\{(\lambda_{0,\epsilon}, \mu^\epsilon) : \epsilon \in (0,1)\}$ is bounded in $\mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^p$, there exists $(\lambda_0, \mu) \in \mathbb{R} \times \mathbb{R}^m \times \mathbb{R}^p$ and subsequences such that $\mu^\epsilon \to \mu = (\mu_1, \mu_2) \in (\mathbb{R}^m \times \mathbb{R}^p)$, and $\lambda_{0,\epsilon} \to \lambda_0 \geq 0$ as $\epsilon \to 0^+$, and we find

$$0 \leq \lambda_0 \left(F_0'(y^*)(\hat{y} - y^*) + F_1(\hat{y}) - F_1(y^*)\right)$$

$$+(\mu_1, G_1'(y^*)(\hat{y} - y^*))_{\mathbb{R}^m} + (\mu_2, G_2(\hat{y}) - G_2(y^*))_{\mathbb{R}^p},$$

for all $\hat{y} \in \mathcal{C}$ with $\hat{y}$ in the effective domain of F_1. Moreover $|(\lambda_0, \mu)| = 1$ and hence (λ_0, μ) is nontrivial.

Since $\mu_2^\epsilon \geq 0$ and $(\mu_2^\epsilon, G_2(y^\epsilon))_{\mathbb{R}^p} \geq 0$, it follows that $\mu_2 \geq 0$ and $(\mu_2, G_2(y^*))_{\mathbb{R}^p} \geq 0$. Moreover $G_2(y^*) \leq 0$, and thus $(\mu_2, G_2(y^*))_{\mathbb{R}^p} = 0$. □

3 Infinite Dimensional Range Case

Let X, Y and Z be a real Banach spaces, and assume that the unit balls in Y^* and Z^* are weakly sequentially compact, which is the case, for example, if X, Y, Z are separable. Further let $K \subset Z$ be a closed, convex cone with vertex at 0, which introduces an ordering on Z such that $u \leq v$ if $u - v \in K$. As in the previous section we consider the minimization problem

$$(P) \qquad \min \quad F(y) \;\text{ subject to } G_1(y) = 0, \;\; G_2(y) \leq 0, \quad y \in \mathcal{C},$$

where $G_1 : X \to Y,\ G_2 : X \to Z$. To state precisely the conditions that will be used we require some preliminaries. The max-operation in $\max(0, G_2(y))$ will be replaced by the distance functional to K given by

$$d_K(\hat{z}) = \inf_{z \in K} |z - \hat{z}|_Z.$$

It is convex and Lipschitz continuous [Cl], pg 50. The convex sub-differential of d_K is defined by

$$\partial d_K(\hat{z}) = \{\xi \in Z^* : d_K(z) - d_K(\hat{z}) \geq \langle \xi, z - \hat{z} \rangle \text{ for all } z \in Z\},$$

where $\hat{z} \in Z$. It was verified in [LY] that $|\xi|_{Z^*} = 1$ for all $\xi \in \partial d_K(\hat{z})$ whenever $\hat{z} \notin K$. Let us note in passing that $\partial d_K(\hat{z})$ is a singleton for any $\hat{z} \notin K$, if Z^* is strictly convex.

Definition 3.1. *$G_2 : X \to Z$ is called d-convex, if*

$$x \to \langle \xi, G_2(x) \rangle_{Z^*,Z} \text{ is convex for all } \xi \in \partial d_K(\hat{z})$$

for all $\hat{z} \in Z$.

To put this condition into context with the common definition of convexity for operators between Banach spaces, let us recall [ET] that $G_2 : X \to Z$ is called convex with respect to the ordering introduced by the cone $K \subset Z$, if

$$G_2(\lambda u + (1-\lambda)v) \le \lambda G_2(u) + (1-\lambda)G_2(v), \tag{3.1}$$

for all $u, v \in K$, and $\lambda \in (0,1)$. Note that, if

$$d_K(\hat{z} + z) \le d_K(\hat{z}) \quad \text{for all } \hat{z} \in Z, \;\; z \in K, \tag{3.2}$$

then

$$\langle \xi, z \rangle_{Z^*,Z} \le 0 \quad \text{for all } z \in K, \;\; \xi \in \partial d_K(\hat{z}), \text{ and } \hat{z} \in Z. \tag{3.3}$$

Thus, if (3.2) holds, then d-convexity in the sense of Definition 3.1 and convexity as in (3.1) coincide. In case $Z = L^p(\Omega)$ or $Z = C(\Omega)$, and $K = \{z : z \le 0 \; a.e.\}$, we have $d_K(G_2(y^\epsilon)) = |\max(0, G_2(y^\epsilon))|$, where the maximum is taken pointwise almost everywhere, and (3.2) holds.

Throughout this section the following assumption is supposed to hold.

$$\text{(H2)} \quad \begin{cases} \text{This is condition (H1) with the requirements on } G \text{ replaced by} \\ G_1 : X \to Y \text{ is } C^1, \\ G_2 : X \to Z \text{ is d-convex and continuous.} \end{cases}$$

In the finite dimensional case (2.10) together with subsequential convergence of $(\lambda_{0,\varepsilon}, \mu^\varepsilon)$ allowed us to argue that the triple $(\lambda_0, \mu_1, \mu_2)$ is nontrivial, and hence guaranteed that the optimality condition (2.1) is qualified. In the finite dimensional case, since the norm is only weakly lower semi-continuous, rather than weakly continuous, an additional condition is required to establish that $(\lambda_0, \mu_1, \mu_2)$ is nontrivial. We use the regular point condition specified in the following theorem.

Theorem 3.1. *Assume that* (H2) *holds with* $Y = \mathbb{R}^m$ *and that the regular point condition*

$$(3.4) \qquad 0 \in int\,\{G_2(y) - G_2(y^*) - K + G_2(y^*),\ y \in \mathcal{C} \cap domF_1\}$$

is satisfied. Let $y^* \in \mathcal{C}$ *be a local minimum of* (P). *Then there exists a nontrivial* $(\lambda_0, \mu_1, \mu_2) \in \mathbb{R}^+ \times Y^* \times Z^*$ *such that*

(3.5)

$$\begin{cases} \lambda_0 \left(F_0'(y^*)(y - y^*) + F_1(y) - F_1(y^*)\right) + \left(\mu_1, G_1'(y^*)(y - y^*)\right)_{\mathbb{R}^m} \\ \qquad\qquad + \langle \mu_2, G_2(y) - G_2(y^*) \rangle_{Z^*,Z} \geq 0 \text{ for all } y \in \mathcal{C} \cap domF_1, \\ \langle \mu_2, z \rangle_{Z^*,Z} \leq 0 \text{ for all } z \in K,\ G_2(y^*) \leq 0,\ \langle \mu_2, G_2(y^*) \rangle_{Z^*,Z} = 0. \end{cases}$$

Proof. We first proceed as in the proof of Theorem 2.1 with the definition of J_ε replaced by

$$J_\epsilon(y) = \left(((F(y) - F(y^*) + \epsilon)^+)^2 + |G_1(y)|_Y^2 + d_K(G_2(y))^2\right)^{\frac{1}{2}},$$

where we do not yet assume that $Y = \mathbb{R}^m$. Again $y_t = y^\varepsilon + t(\hat{y} - y^\varepsilon)$, $t \in (0, 1)$, with $\hat{y} \in \mathcal{C} \cap\ domF_1$.

We first assume that there exists a subsequence of ε, denoted by the same symbol, such that $G_2(y^\varepsilon) \notin K$ for all ε sufficiently small. The converse will be considered further below.

The first term on the right hand side of (2.5) is estimated as before. To estimate the second term we first note that

$$(3.6) \qquad |G_1(y_t)|_Y^2 - |G_1(y^\epsilon)|_Y^2 \leq\ 2\,\langle \tilde{\mu}_1^{\epsilon,t}, G_1(y_t) - G_1(y^\epsilon) \rangle_{Y^*,Y}\,,$$

where $\tilde{\mu}_1^{\epsilon,t} \in \partial\varphi(G_1(y_t))$ and $\varphi(v) = \frac{|v|_Y^2}{2}$. Since $\partial\varphi$ coincides with the duality mapping $F : Y \to Y^*$, see e.g. [M], pg 44, we have

$$|\tilde{\mu}_1^{\epsilon,t}|_{Y^*} = |G_1(y_t)|_Y.$$

In particular for each fixed $\epsilon > 0$ the family $\{\tilde{\mu}_1^{\epsilon,t} : t \in (0, 1)\}$ is bounded in Y^*. Since the unit ball in Y^* is assumed to be weakly sequentially compact, there exists a subsequence, denoted by the same symbol, that converges weakly* to some $\tilde{\mu}_1^\epsilon \in Y^*$ as $t \to 0^+$. Passing to the limit in

$$\varphi(G_1(y_t)) + \langle \tilde{\mu}_1^{\epsilon,t}, v - G_1(y_t) \rangle_{Y^*,Y} \leq \varphi(v), \quad \text{for } v \in Y,$$

we find

$$\varphi(G_1(y^\epsilon)) + \langle\, \tilde\mu_1^\epsilon, v - G_1(y^\epsilon)\,\rangle_{Y^*,Y} \le \varphi(v), \qquad \text{for } v \in Y,$$

and hence $\tilde\mu_1^\epsilon \in \partial\varphi(G_1(y^\epsilon))$ and $|\tilde\mu_1^\epsilon|_{Y^*} = |G_1(y^\epsilon)|_Y$. For $t \to 0^*$ we have the estimate

$$\langle\, \tilde\mu_1^{\epsilon,t}, G_1(y_t) - G_1(y^\epsilon)\rangle - t\langle \tilde\mu_1^\varepsilon, G_1'(y^\epsilon)(\hat y - y^\epsilon)\rangle$$
$$= \langle\, \tilde\mu_1^{\epsilon,t}, G_1(y_t) - G_1(y^\epsilon) - tG_1'(y^\epsilon)(\hat y - y^\epsilon)\, +\, t\,\langle \tilde\mu_1^{\epsilon,t} - \tilde\mu_1^\epsilon, G_1'(y^\epsilon)(\hat y - y^\epsilon)\rangle = o(|t|),$$

where the duality products are taken from Y to Y^*. Combined with (3.6) this implies

$$|G_1(y_t)|_Y^2 - |G_1(y^\epsilon)|_Y^2 \le\; 2\,t\,\langle\, \tilde\mu_1^\epsilon, G'(y^\epsilon)(\hat y - y^\epsilon)\,\rangle_{Y^*,Y} + o(|t|), \tag{3.7}$$

where $\tilde\mu_1^\epsilon \in \partial\varphi(G_1(y^\epsilon))$.

For the third term we obtain

$$\begin{aligned} &d_K(G_2(y_t))^2 - d_K(G_2(y^\varepsilon))^2 \\ &\quad = (d_K(G_2(y_t)) + d_K(G_2(y^\varepsilon)))\,(d_K(G_2(y_t)) - d_K(G_2(y^\varepsilon))) \\ &\quad \le (d_K(G_2(y_t)) + d_K(G_2(y^\varepsilon)))\,\langle \xi_t, G_2(y_t) - G_2(y^\varepsilon)\rangle_{Z^*,Z} \\ &\quad \le t(d_K(G_2(y_t)) + d_K(G_2(y^\varepsilon)))\,\langle \xi_t, G_2(\hat y) - G_2(y^\varepsilon)\rangle_{Z^*,Z}, \end{aligned}$$

where $\xi_t \in \partial_K(G_2(y_t))$ and in the last estimate we used d-convexity of G_2. Setting

$$\tilde\mu_2^{\varepsilon,t} = (d_K(G_2(y_t)) + d_K(G_2(y^\varepsilon)))\,\xi_t, \quad \text{with } \xi_t \in \partial d_K(G_2(y_t)), \tag{3.8}$$

we thus have

$$d_K(G_2(y_t))^2 - d_K(G_2(y^\varepsilon))^2 \le t\,\langle \tilde\mu_2^{\epsilon,t}, G_2(\hat y) - G_2(y^\epsilon)_{Z^*,Z}. \tag{3.9}$$

As in the proof of Theorem 2.1 we set

$$\tilde\alpha^{\epsilon,t} = ((F(y_t) - F(y^*) + \epsilon)^+ + (F(y^\epsilon) - F(y^*) + \epsilon)^+) \text{ and } \tilde\mu_1^\varepsilon = 2G_1(y^\varepsilon).$$

Combining (2.5) with $\max(0, G_2)$ replaced by $d_K(G_2)$, (2.8), (3.7) and (3.9) we arrive at

(3.10)

$$\begin{aligned} &-\sqrt{\epsilon}\, d(y_t, y^\epsilon) \\ &\quad \le \tfrac{1}{J_\epsilon(y^\epsilon) + J_\epsilon(y_t)}\, \big[\tilde\alpha^{\epsilon,t}\big(tF_0'(y^\epsilon)(\hat y - y^\epsilon) + t(F_1(\hat y) - F_1(y^\epsilon))\big) \\ &\qquad +2\,t\,\langle \tilde\mu_1^\epsilon, G_1'(y^\epsilon)(\hat y - y^\epsilon)\rangle_{Y^*,Y} + t\,\langle \tilde\mu_2^{\epsilon,t}, G_2(\hat y) - G_2(y^\epsilon)\rangle_{Z^*,Z}\big] + o\,(|t|), \end{aligned}$$

for all $t \in (0, \bar{t}(\epsilon))$. Let

$$\alpha^{\epsilon,t} = \frac{\tilde{\alpha}^{\epsilon,t}}{J_\epsilon(y_t) + J_\epsilon(y^\epsilon)}, \quad {\mu_1}^{\epsilon,t} = \frac{2\tilde{\mu}_1^\epsilon}{J_\epsilon(y_t) + J_\epsilon(y^\epsilon)}, \quad \mu_2^{\epsilon,t} = \frac{\tilde{\mu}_2^{\epsilon,t}}{J_\epsilon(y_t) + J_\epsilon(y^\epsilon)}.$$

Taking the limit as $t \to 0^+$ we have

$$\lambda_{0,\epsilon} := \lim_{t\to 0^+} \alpha^{\epsilon,t} = \tfrac{(F(y^\epsilon) - F(y^*) + \epsilon)^+}{J_\epsilon(y^\epsilon)}, \quad \mu_1^\epsilon := \lim_{t\to 0^+} \mu_1^{\epsilon,t} = \tfrac{\tilde{\mu}_1^\epsilon}{J_\epsilon(y^\epsilon)}.$$

Since $G_2(y^\varepsilon)$ is not in K by assumption and since K is assumed to be closed, it follows that $G_2(y_t) \notin K$ for all $t \geq$ sufficiently small. As a consequence $|\xi_t|_{Z^*} = 1$ for all t sufficiently small. Since by assumption the unit sphere in Z^* is weakly sequentially compact, there exists a subsequence that converges weakly-star to some element $\xi^\varepsilon \in \partial d_K(G_2(y_\varepsilon))$ as $t \to 0^+$. We have

$$\mu_2^\varepsilon := \mathrm{w}^* - \lim_{t\to 0^+} \mu_2^{\varepsilon,t} = \frac{d_K(G_2(y^\varepsilon))\, \xi^\varepsilon}{J_\varepsilon(y^\varepsilon)}. \tag{3.11}$$

Since $|\xi_\varepsilon|_{Z^*} = 1$ as a consequence of $\xi_\varepsilon \in \partial d_K(G(y_\varepsilon))$, and since $|\tilde{\mu}_1^\varepsilon|_{Y^*} = |G_1(y^\varepsilon)|_Y$, we have

$$|(\lambda_{0,\varepsilon}, \mu^\epsilon)|_{\mathbb{R}\times Y^* \times Z^*} = 1. \tag{3.12}$$

Moreover, $\lambda_{0,\epsilon} \geq 0$. Dividing (3.10) by t and letting $t \to 0^+$, we obtain

$$\begin{aligned} -\sqrt{\epsilon} d(\hat{y}, y^\epsilon) \leq{} & \lambda_{0,\epsilon} \left(F_0'(y^\epsilon)(\hat{y} - y^\epsilon) + F_1(\hat{y}) - F_1(y^\epsilon)\right) \\ & + \langle \mu_1^\epsilon, G_1'(y^\epsilon)(\hat{y} - y^\epsilon)\rangle_{Y^*,Y} + \langle \mu_2^\epsilon, G_2(\hat{y}) - G_2(y^\epsilon)\rangle_{Z^*,Z}. \end{aligned} \tag{3.13}$$

Since $\{(\lambda_{0,\epsilon}, \mu_1^\epsilon, \mu_2^\epsilon) : \epsilon \in (0,1)\}$ is bounded in $\mathbb{R} \times Y^* \times Z^*$, there exists $(\lambda_0, \mu_1, \mu_2) \in \mathbb{R} \times Y^* \times Z^*$ and subsequences, denoted by the same symbol, such that $\lambda_{0,\epsilon} \to \lambda_0 \geq 0$, $(\mu_1^\epsilon, \mu_2^\epsilon) \to (\mu_1, \mu_2)$ weakly*, as $\epsilon \to 0^+$, and we find

$$\begin{aligned} 0 \leq{} & \lambda_0 \left(F_0'(y^*)(\hat{y} - y^*) + F_1(\hat{y}) - F_1(y^*)\right) \\ & + \langle \mu_1, G_1'(y^*)(\hat{y} - y^*)\rangle_{Y^*,Y} + \langle \mu_2, G_2(\hat{y}) - G_2(y^*)\rangle_{Z^*,Z}, \end{aligned} \tag{3.14}$$

for all $\hat{y} \in \mathcal{C}$ with $\hat{y}$ in the effective domain of F_1.

To argue complementarity, note that

$$\langle \xi_\varepsilon, y - G_2(y^\varepsilon)\rangle \leq -d_K(G_2(y^\varepsilon), \text{ for all } y \in K,$$

since $\xi_\varepsilon \in \partial d_K(G_2(y^\varepsilon))$, and hence

$$\langle \mu_2^\varepsilon, y - G_2(y^\varepsilon)\rangle \le 0, \text{ for all } y \in K.$$

Taking the limit $\varepsilon \to 0$ we find

$$\langle \mu_2, y - G_2(y^*)\rangle_{Z^*,Z} \le 0, \text{ for all } y \in K. \tag{3.15}$$

Since K is a convex cone, we have $y + G_2(y^*) \in K$ for any $y \in K$. This implies that $\langle \mu_2, y\rangle \le 0$, for all $y \in K$ and in particular $\langle \mu_2, G_2(y^*)\rangle \le 0$. Setting $y = 0$ in (3.15) we have $\langle \mu_2, G_2(y^*)\rangle \ge 0$ and hence $\langle \mu_2, G_2(y^*)\rangle = 0$, as desired.

Now we consider the case that $G_2(y^\varepsilon) \in K$ for all ε sufficiently small. The $|\xi_t|_{Z^*} \le 1$ for all $t \ge$. By (3.8) and (3.11) it follows that $\mu_2^\varepsilon = 0$ for all $\varepsilon > 0$ sufficiently small. Consequently $\mu^2 = 0$ as well, and we can follow the above steps to argue that again (3.5) holds.

It remains to argue non-triviality of $(\lambda_0, \mu_1, \mu_2)$ in (3.14). Henceforth we assume that $Y = \mathbb{R}^m$. Then the subsequence that we chose from μ_1^ε converges strongly to μ_1.

Assume that $(\lambda_{0,\epsilon}, \mu_1^\varepsilon) \to 0^+$ as $\epsilon \to 0$. We shall show that (3.4) implies that μ_2 is nontrivial. First note that as a consequence of (3.12)

$$\lim_{\varepsilon\to 0}(\ \lambda_{0,\varepsilon}^2 + |\mu_1^\varepsilon|_{\mathbb{R}^m}^2 + |\mu_2^\varepsilon|_{Z^*}^2\ ) = \lim_{\varepsilon\to 0} |\mu_2^\varepsilon|_{Z^*}^2 = 1. \tag{3.16}$$

Since $\xi^\varepsilon \in \partial d_K(G_2(y_\varepsilon))$ we have $\langle \xi^\varepsilon, y - G_2(y^\varepsilon)\rangle_{Z^*,Z} \le 0$ for all $y \in K$, and hence

$$\langle \mu_2^\varepsilon, y - G_2(y^\varepsilon)\rangle_{Z^*,Z} \le 0 \text{ for all } y \in K. \tag{3.17}$$

From (3.13) we deduce that

$$-O(\varepsilon) \le \langle \mu_2^\varepsilon, G_2(\hat{y}) - G_2(y^\varepsilon)\rangle_{Z^*,Z},$$

where $O(\varepsilon)$ denotes a quantity that converges to 0 as $\varepsilon \to 0^+$ and $\hat{y} \in \mathcal{C} \cap domF_1$. Combining these two statements we find that

$$-O(\varepsilon) \le \langle \mu_2^\varepsilon, G_2(\hat{y}) - y)\rangle_{Z^*,Z} \le 0 \text{ for all } y \in K. \tag{3.18}$$

From (3.4) it follows that there exists $0 \ne z_0 \in Z$ and $\rho > 0$ such that

$$G_2(\hat{y}) - y = z_0 + \eta$$

has a solution $(\hat{y}, y) \in (\mathcal{C} \cap domF_1) \times K$ for all η in the ball $B_Y(0,\rho)$. Hence

$$-\langle \mu_2^\varepsilon, \eta \rangle_{Z^*,Z} \leq \langle \mu_2^\varepsilon, z_0 \rangle_{Z^*,Z} + O(\varepsilon) \text{ for all } \eta \in B_Y(0,\rho).$$

Taking the supremum of the left hand side over $\eta \in B_Y(0,\rho)$, we obtain

$$\rho |\mu_2^\varepsilon|_{Z^*} \leq \langle \mu_2^\varepsilon, z_0 \rangle_{Z^*,Z} + O(\epsilon).$$

By (3.16) we obtain

$$\rho \leq \langle \mu_2, z_0 \rangle_{Z^*,Z}.$$

Hence $\mu_2 = 0$ is impossible. □

As a consequence of the proof we find the following corollary which provides modifications to assumption (3.4).

Corollary 3.1. *If instead of* (3.4)*, the cone* K *contains an interior point, or* $0 \in int\{G_2(y) - G_2(y^*) : y \in \mathcal{C} \cap domF_1\}$*, the conclusion of Theorem 3.1 remains correct.*

Proof. We proceed as in the proof of Theorem 3.1. Choosing $\hat{y} = y^*$ in (3.18) we find

$$-O(\varepsilon) \leq \langle \mu_2^\varepsilon, G_2(y^*) - y \rangle_{Y^*,Y} \text{ for all } y \in K.$$

If $intK \neq \emptyset$ there exists a ball $B(z_0,\rho) \subset K$, where z_0 can be chosen differently from $G_2(y^*)$. Consequently

$$-\langle \mu_2^\varepsilon, \eta \rangle_{Z^*,Z} \leq \langle \mu_2^\varepsilon, G_2(y^*) - z_0 \rangle_{Z^*,Z} + O(\varepsilon), \text{ for all } \eta \in B(0,\rho),$$

and we can argue as in the proof of Theorem 3.1 that $\mu_2 \neq 0$.

Turning to the case $0 \in \{G_2(\mathcal{C} \cap domF_1) - G_2(y^*)\}$, note first that by (3.17) and (2.2)

$$\langle \mu_2^\varepsilon, y - G_2(y^*) \rangle_{Z^*,Z} \leq \langle \mu_2^\varepsilon, G_2(y^\varepsilon) - G_2(y^*) \rangle_{Z^*,Z}, \text{ for all } y \in K.$$

Using $0 \in \{G_2(\mathcal{C} \cap domF_1) - G_2(y^*)\}$ we can argue that there exists $z_0 \neq 0$ and $\rho > 0$ such that

$$\langle \mu_2^\varepsilon, \eta - z_0 \rangle_{Z^*,Z} \leq O(\varepsilon), \text{ for all } \eta \in B(0,\rho),$$

which allows to conclude that $\mu_2 \neq 0$. □

Theorem 3.2. *Assume that* (H2) *holds with* $Z = \mathbb{R}^p$ *and that the regular point condition*

$$0 \in int\,\{G_1'(y^*)(y - y^*),\ \ y \in \mathcal{C} \cap domF_1\}. \tag{3.19}$$

is satisfied. Then the conclusion of Theorem 3.1 remains correct.

Proof. In view of the proof to the previous theorem, we only need to verify that $(\lambda_0, \mu_1, \mu_2)$ is nontrivial. Assume that $(\lambda_{0,\varepsilon}, \mu_2^\varepsilon) \to 0$ as $\varepsilon \to 0^+$. We shall prove that $\mu_1 \in Y^*$ is nontrivial. As a consequence of (3.12), we have $\lim_{\varepsilon \to 0^+} |\mu_1^\varepsilon|_{Y^*} = 1$. By (3.19) there exists $0 \neq y_0 \in X$ and $\rho > 0$ such that for any $\eta \in B_Y(0, \rho)$ there exists $\hat{y} \in \mathcal{C}$ such that

$$G_1'(y^*)(\hat{y} - y^*) = y_0 + \eta.$$

Thus

$$\langle \mu_1^\epsilon, G_1'(y^\epsilon)(\hat{y} - y^*)\rangle_{Y^*,Y} = \langle \mu_1^\epsilon, y_0 + \eta\rangle_{Y^*,Y} + \langle \mu_1^\epsilon, (G_1'(y^\epsilon) - G_1'(y^*))(\hat{y} - y^*)\,\rangle_{Y^*,Y},$$

where

$$|\langle \mu_1^\epsilon, (G_1'(y^\epsilon) - G_1'(y^*))(\hat{y} - y^*)\rangle_{Y^*,Y}| \to 0 \text{ as } \epsilon \to 0^+.$$

It follows from (3.13) that

$$-\langle \mu_1^\epsilon, \eta\rangle_{Y^*,Y} \leq \langle \mu_1^\epsilon, y_0\rangle_{Y^*,Y} + O(\epsilon)$$

where $O(\epsilon) \to 0$ as $\epsilon \to 0^+$. Taking the supremum of the left hand side over $\eta \in B_Y(0, \rho)$, we obtain

$$\rho |\mu_1^\epsilon|_{Y^*} \leq \langle \mu_1^\epsilon, y_0\rangle_{Y^*,Y} + O(\epsilon).$$

Since $\lim_{\epsilon \to 0^+} |\mu_1^\epsilon|_{Y^*} \to 1$, letting $\epsilon \to 0^+$, we obtain

$$\rho \leq \langle \mu_1, y_0\rangle_{Y^*,Y}.$$

Hence $\mu_1 = 0$ is impossible. □

Combining the regular point conditions of Theorems 3.1 and 3.2 results in the condition

$$0 \in \text{int} \begin{pmatrix} \{\, G_1'(y^*)(y - y^*) : y \in \mathcal{C} \cap domF_1\} \\ \{\, G_2(y) - k : \mathcal{C} \cap domF_1, k \in K\} \end{pmatrix}, \tag{3.20}$$

where the right hand side is a subset of $Y \times Z$. In the following theorem we shall show that (3.20) implies that $(\lambda_0, \mu_1, \mu_2)$ is nontrivial without requiring finite dimensionality of the range spaces of either G_1 or G_2. This is a simple consequence of the proofs of Theorems 3.1 and 3.2. Moreover, (3.20) is a sufficient condition for normality, i.e. $\lambda_0 \neq 0$.

Theorem 3.3. *Assume that* (H2) *and* (3.20) *hold. Then the conclusion of Theorem 3.1 remains correct. Moreover the solution y^* to (P) is normal, i.e. the necessary condition* (3.5) *holds with $\lambda_0 = 1$.*

Proof. If $\lambda_{0,\varepsilon} \to 0$ then by (3.13)

$$-O(\varepsilon) \leq \langle\, \mu_1^\varepsilon, G_1'(y^\varepsilon)(\hat{y} - y^\varepsilon)\,\rangle_{Y^*,Y} + \langle\, \mu_2^\varepsilon, G_2(\hat{y}) - G_2(y^\varepsilon)\,\rangle_{Z^*,Z}.$$

Using (3.17)

$$-O(\varepsilon) \leq \langle\, \mu_1^\varepsilon, G_1'(y^\varepsilon)(\hat{y} - y^\varepsilon)\,\rangle_{Y^*,Y} + \langle\, \mu_2^\varepsilon, G_2(\hat{y}) - y\,\rangle_{Z^*,Z} \text{ for all } y \in K.$$

The regular point condition (3.20) implies the existence of $(y_0, z_0) \in Y \times Z$, both nonzero and $\rho > 0$, such that

$$0 \leq \langle\, \mu_1^\varepsilon, \eta_1 + y_0\,\rangle_{Y^*,Y} + \langle\, \mu_2^\varepsilon, \eta_2 + z_0\,\rangle_{Z^*,Z} + O(\varepsilon), \text{ for all } (\eta_1, \eta_2) \in B_{Y\times Z}\big(0, \rho\big).$$

This implies that

$$\rho|(\mu_1^\varepsilon, \mu_2^\varepsilon)|_{Y^*,Z^*} \leq \langle\, \mu_1^\varepsilon, y_0\,\rangle_{Y^*,Y} + \langle\, \mu_2^\varepsilon, z_0\,\rangle_{Z^*,Z}, + O(\varepsilon).$$

Passing to the limit we find, using (3.12),

$$\rho \leq \langle\, \mu_1, y_0\,\rangle_{Y^*,Y} + \langle\, \mu_2, z_0\,\rangle_{Z^*,Z},$$

and hence μ_1, μ_2 cannot both be 0.

To verify the second assertion of the theorem we once again use the regular point condition (3.20). Hence for all $(\hat{\mu}_1, \hat{\mu}_2)$ belonging to a neighborhood of 0 in $Y^* \times Z^*$, there exist elements $y \in \mathcal{C} \cap domF_1$, $k \in K$ such that

$$(\hat{\mu}_1, \hat{\mu}_2) = (G_1'(y^*)(y - y^*), G_2(y) - G_2(y^*) - k + G_2(y^*)).$$

Consequently

$$\begin{aligned}&\langle\, \mu_1, \hat{\mu}_1\rangle_{Y^*,Y} + \langle \mu_2, \hat{\mu}_2\rangle_{Z^*,Z} \\ &\quad = \langle\, \mu_1, G_1'(y^*)(y - y^*)\,\rangle_{Y^*,Y} + \langle \mu_2,\, G_2(y) - G_2(y^*) - k + G_2(y^*)\,\rangle_{Z^*,Z}.\end{aligned}$$

Note that $\langle \mu_2, k - G_2(y^*)\rangle = \langle \mu_2, k\rangle \le 0$ for $k \in K$. If $\lambda_0 = 0$, then the first equation in (3.5) implies that

$$\langle \mu_1, \hat{\mu}_1\rangle_{Y^*,Y} + \langle \mu_2, \hat{\mu}_2\rangle_{Z^*,Z} \ge \langle\, \mu_1, G_1'(y^*)(y-y^*)\,\rangle_{Y^*,Y} + \langle\, \mu_2, G_2(y) - G_2(y^*)\,\rangle_{Z^*,Z} \ge 0$$

for all $(\hat{\mu}_1, \hat{\mu}_2)$ in a neighborhood of 0 and thus $\mu_1 = \mu_2 = 0$, which is a contradiction. Consequently $\lambda_0 > 0$ and thus the problem is strictly normal. By rescaling $(\lambda_0, \mu_1, \mu_2)$ one can set $\lambda^0 = 1$. $\square$

4 L^1-minimum Norm Control

Consider the optimal exit problem with minimum L^1- norm

$$(4.1) \qquad \begin{cases} \min_{u,\tau} \quad \int_0^\tau (f(x(t)) + \delta|u(t)|)\,dt \quad \text{subject to} \\ \frac{d}{dt}x = b(x(t), u(t)), \quad x(0) = x_0, \\ g(x(\tau) = 0, \quad |u(t)|_{\mathbb{R}^k} \le \gamma \text{ for a.e. } t, \end{cases}$$

where $\delta > 0$, $x_0 \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, $b : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}^m$ are smooth functions. We have two motivations to consider (4.1). In the context of sparse controls the pointwise norm constraints, allow us to avoid controls in measure space. In fact, an L^1- cost for the control in the cost does not guarantee existence of a minimizer in this space. In the case of time optimal control problems the term $\delta \int_0^T |u|\,dt$ can be considered as regularization term. We shall see from the optimality condition (4.9) below that this determines the control as a function of the adjoint by means of an equation rather than an inclusion which would be the case if no regularization was used.

One can transform (4.1) to the fixed interval $s \in [0, 1]$ via the change of variable $t = \tau\, s$ leading to

$$(4.2) \qquad \begin{cases} \min_{u,\tau} \quad \int_0^1 \tau\left(f(x(t)) + \delta|u(t)|\right)dt \quad \text{subject to} \\ \frac{d}{dt}x(t) = \tau\, b(x(t), u(t)), \quad x(0) = x_0, \\ g(x(1)) = 0, \quad u \in U_{ad} = \{u \in L^\infty((0,1); \mathbb{R}^k) : |u(t)| \le \gamma\}. \end{cases}$$

In terms of the notation set forth in Section 2, set $y = (u, \tau)$ and define

$$F_0(y) = \tau \int_0^1 f(x(t))\,dt, \quad F_1(u) = \delta \int_0^1 |u(t)|\,dt,$$
$$F(y) = F_0(y) + \tau\, F_1(u), \quad G(y) = g(x(1)),$$

where $x = x(\cdot; u, \tau)$ is the solution to the initial value problem in (4.2), given $u \in U_{ad}$ and $\tau \geq 0$. In the context of the general framework we set $X = L^2((0,1); \mathbb{R}^k) \times \mathbb{R}$, $\mathcal{C} = U_{ad}$, $G = G_1$. Note that τ was not incorporated to the definition of F_1, this would destroy its convex structure. The appearance of the multiplying factor τ will require us to slightly extend the general theory of Section 3, to obtain a necessary optimality condition for (4.2).

The control problem can now be equivalently formulated as

$$\min_{(u,\tau) \in U_{ad} \times \mathbb{R}^+} F(y) \quad \text{subject to } G(y) = 0. \tag{4.3}$$

We assume that $y^* = (u^*, \tau^*)$ is an optimal solution to (4.3) with $\tau^* > 0$.

We impose that the regular point condition

$$0 \in int \, \{G_u(y^*)(v - u^*) + G_\tau(y^*)(\tau - \tau^*) : v \in U_{ad}, \tau > 0\} \tag{4.4}$$

holds at $y^* = (u^*, \tau^*)$. We now extend the proof of Theorem 3.1 by replacing the expression $F(y_t) - F(y^\epsilon)$ in (2.8) by

$$F_0(y_t) - F_0(y^\epsilon) + \tau_t F_1(u_t) - \tau^\varepsilon F_1(u^\varepsilon),$$

where $\tau_t = \tau^\varepsilon + t(\hat{\tau} - \tau^\varepsilon),\ u_t = u^\varepsilon + t(\hat{u} - u^\varepsilon)$. Noting that

$$\tau_t F_1(u_t) - \tau^\varepsilon F_1(u^\varepsilon) = t(\hat{\tau} - \tau^\varepsilon) F_1(u_t) + \tau^\varepsilon (F_1(u_t) - F_1(u\varepsilon)))$$

the following steps can be carried out as before. We find that there exist a Lagrange multiplier $(\lambda_0, \mu) \in \mathbb{R}^+ \times \mathbb{R}^m$ such that

$$\lambda_0[\tau^* \left(F_1(u) - F_1(u^*)\right) + (\tau - \tau^*) \, F_1(u^*) + (F_0(y^*))_u (u - u^*) + (F_0(y^*))_\tau (\tau - \tau^*)]$$
$$+ \mu^T G_u(y^*)(u - u^*) + + \mu^T G_\tau(y^*)(\tau - \tau^*) \geq 0,$$

for all $u \in U_{ad}$ and $\tau \geq 0$.

As in the proof of Theorem 3.3 the regular point condition can be used to argue that $\lambda_0 > 0$ and hence by rescaling μ it can be chosen to be 1. We arrive at

(4.5)

$$\tau^* \left(F_1(u) - F_1(u^*)\right) + (\tau - \tau^*) \, F_1(u^*)$$
$$+ \big((F_0(y^*))_u + \mu^T G_u(y^*)\big)(u - u^*) + \big((F_0(y^*))_\tau + \mu^T G_\tau(y^*)\big)(\tau - \tau^*) \geq 0,$$

for all $u \in U_{ad}$ and $\tau \geq 0$. Note that for $v \in L^\infty((0,1);\mathbb{R}^k))$

$$G_u(y^*)(v) = g_x(x^*(1))\,h(1), \quad G_\tau(y^*) = g_x(x^*(1))\xi(1),$$

$$(F_0(y^*))_u(v) = \tau^* \int_0^1 (f'(x^*(t)), h(t))_{\mathbb{R}^n}\,dt,$$

$$(F_0(y^*))_\tau(v) = \int_0^1 \big((\tau^*\, f'(x^*(t)), \xi(t))_{\mathbb{R}^n} + f(x^*(t))\big)\,dt,$$

where (h,ξ) satisfies

$$\begin{aligned} &\frac{d}{dt}h(t) = \tau^* \big(b_x(x^*(t),u^*(t))h(t) + b_u(x^*(t),u^*(t))v(t)\big),\ h(0)=0, \\ &\frac{d}{dt}\xi(t) = \tau^*\, b_x(x^*(t),u^*(t))\xi(t) + b(x^*(t),u^*(t), \quad \xi(0)=0. \end{aligned} \tag{4.6}$$

Let $p \in H^1((0,1);\mathbb{R}^n)$ satisfy the adjoint equation

$$-\frac{d}{dt}p(t) = \tau^* \big(b_x(x^*(t),u^*(t))^T p(t) + f'(x^*(t))\big), \quad p(1)^T = \mu^T g_x(x^*(1)). \tag{4.7}$$

Then

$$\begin{aligned} (h(1),p(1))_{\mathbb{R}^n} &= \int_0^1 \frac{d}{dt}(h(t),p(t))_{\mathbb{R}^n} \\ &= \tau^* \int_0^1 \big((b_u(x^*(t),u^*(t))v(t),p(t))_{\mathbb{R}^n} - (f'(x^*(t)),h(t))\big)_{\mathbb{R}^n}\,dt \\ (\xi(1),p(1))_{\mathbb{R}^n} &= \int_0^1 \frac{d}{dt}(\xi(t),p(t))_{\mathbb{R}^n} \\ &= \int_0^1 \big((b(x^*(t),u^*(t)),p(t))_{\mathbb{R}^n} - \tau^*(f'(x^*(t)),\xi(t))_{\mathbb{R}^n}\big)\,dt. \end{aligned}$$

Using these equalities and $p(1)^T = \mu^T g_x(x^*(1))$ we have

$$\begin{aligned} &\big(F_0(y^*)_u + \mu^T G(y^*)_u\big)(u-u^*) + \big(F_0(y^*)_\tau + \mu^T G(y^*)_\tau\big)(\tau-\tau^*) \\ &= \tau^* \textstyle\int_0^1 (f'(x^*(t)),h(t))_{\mathbb{R}^n}\,dt + p(1)^T h(1) + (\tau-\tau^*)p(1)^T\xi(1) \\ &\qquad (\tau-\tau^*) \textstyle\int_0^1 \big(\tau^*(f'(x^*(t)),\xi(t))_{\mathbb{R}^n} + f(x^*(t))\big)\,dt \\ &\qquad = \tau^* \textstyle\int_0^1 (b_u(x^*(t),u^*(t))\,(u(t)-u^*(t)),p(t))_{\mathbb{R}^n}\,dt + (\tau-\tau^*) \int_0^1 \big(b(x^*(t),u^*(t)),p(t)\big)_{\mathbb{R}^n}\,dt \\ &\qquad +(\tau-\tau^*) \textstyle\int_0^1 (f(x^*(t))\,dt. \end{aligned}$$

From (4.5) therefore we find for all $u \in U_{ad}$ and $\tau \geq 0$

(4.8)

$$(\tau - \tau^*) \int_0^1 \big(f(x^*(t)) + \delta|u^*(t)| + (b(x^*(t), u^*(t)), p(t))_{\mathbb{R}^n}\big)\, dt$$

$$+\tau^* \int_0^1 \big((b_u(x^*(t), u^*(t))^t p(t), u(t) - u^*(t))_{\mathbb{R}^k} + \delta|u(t)| - \delta|u^*(t)|)\big)\, dt \geq 0.$$

Setting $\tau = \tau^*$ we obtain the optimality condition

(4.9)

$$u^*(t) = \begin{cases} 0 & \text{if } |b_u(x^*(t), u^*(t))^T p(t)| < \gamma \\ -\gamma \dfrac{b_u(x^*(t), u^*(t))^T p(t)}{|b_u(x^*(t), u^*(t))^T p(t)|} & \text{if } |b_u(x^*(t), u^*(t))^T p(t)| \geq \gamma. \end{cases}$$

Here we note that pointwise the integrand in the second expression on the left hand side of (4.8) is the necessary optimality condition for

$$\min_{|v|_{\mathbb{R}^k} \leq \gamma} (b_u(x^*(t), u^*(t))^t p(t), v)_{\mathbb{R}^k} + \delta |v|_{\mathbb{R}^k}.$$

Setting $u = u^*$ in (4.8) we find

$$\int_0^1 \big(f(x^*(t)) + \delta|u^*(t)| + (b(x^*(t), u^*(t)), p(t))_{\mathbb{R}^n}\big)\, dt = 0. \tag{4.10}$$

The Hamiltonian associated to (4.2) is given by

$$\mathcal{H}(t) = f(x(t)) + \delta|u(t)| + \big(b(x(t), u(t)), p(t)\big)_{\mathbb{R}^n},$$

is constant along the optimal trajectory x^* and control u^*. For a proof of this fact we refer to the Appendix. Together with (4.10) this implies that

$$\mathcal{H}(t) = f(x^*(t)) + |u^*(t)| + (b(x^*(t), u^*(t)), p(t))_{\mathbb{R}^n} = 0 \text{ on } [0, 1]. \tag{4.11}$$

Theorem 4.1. *Suppose* $(u^*, \tau^*) \in U_{ad} \times R^+$ *is optimal for* (4.2) *with* $\tau^* > 0$ *and let the regular point condition* (4.4) *holds. Then, there exist a Lagrange multiplier* $\mu \in \mathbb{R}^m$ *such that the optimality conditions* (4.9), (4.11) *hold, where* $p \in H^1((0, 1); R^n)$ *satisfies the adjoint equation* (4.7).

The regular point condition holds if the linearized non-autonomous control system

$$(4.12)\quad \begin{cases} \frac{d}{dt}h(t) = \tau^*(b_x(x^*(t),u^*(t))h(t) + b_u(x^*(t),u^*(t))(u(t)-u^*(t))), \\ h(0)=0, \quad \text{with } u \in U_{ad}, \end{cases}$$

is controllable with constraints on the controls , i.e., $0 \in int\, R(1)$ where $R(1) = \{h(1;u) : u \in U_{ad}\}$. For example, if $S = \{t \in (0,1) : u^*(t) = 0\}$ contains an open interval and $b(x,u) = Ax + Bu$ is a linear control system, such that (A,B) is controllable, then we have that $0 \in int\, R(1)$.

5 L^∞-norm Minimization

In this short section we consider the L^∞-norm minimization problem

$$(5.1)\qquad \min_{y\in X_0} \quad |\Lambda y|_{L^\infty} + F_0(y) \text{ subject to } G_1(y) = 0,$$

where $F_0 \in C^1(X_0,\mathbb{R})$, $G_1 \in C^1(X_0,Y)$, $\Lambda \in \mathcal{L}(X_0,L^2(\Omega))$, with X_0 and Y real Banach spaces, and the unit ball in Y^* weakly sequentially compact. Problem (5.1) can be equivalently be expressed as

$$(5.2)\qquad \begin{aligned} &\min_{\gamma\in\mathbb{R}^+, y\in X_0} \quad \gamma + F_0(y) \\ &\text{subject to } G_1(y) = 0, \text{ and } |\Lambda y|_{L^\infty} \le \gamma. \end{aligned}$$

To avoid complications with a constraint set that depends on the parameter γ a parametrization according to $y = \gamma z$ is performed and (5.2) is transformed into

$$(5.3)\qquad \begin{cases} \min_{\gamma\in\mathbb{R}^+, z\in X_0} \quad \gamma + F_0(\gamma z) \\ \text{subject to } G_1(\gamma z) = 0, \quad z \in \mathcal{C}, \end{cases}$$

where $\mathcal{C} = \{z \in X_0 : |\Lambda z|_{L^\infty} \le 1\}$. We set

$$x = (\gamma, y) \in X = \mathbb{R} \times X_0, \quad F(\gamma,y) = \gamma + F_0(y).$$

In this way, (5.3) is a special case of (P). We suppose that (5.3) admits a solution γ^*, z^* and set $y^* = \gamma^* z^*$. Further the regular point condition

$$0 \in \{G_1'(\gamma^* z^*)(\gamma^* v - \gamma^* z^*) + G_1'(\gamma^* z^*)(\gamma - \gamma^*)z^* : v \in \mathcal{C}, \gamma \ge 0\}.$$

is supposed to hold. Setting $\gamma = 2\gamma^*$ it is implied by $0 \in \{G_1'(\gamma^* z^*)v : v \in \mathcal{C}\}$.

Then by Theorem 3.3 there exists $\mu \in Y^*$ such that

$$\begin{cases} \langle F_0'(\gamma^* z^*) + G_1'(\gamma^* z^*)^* \mu, \phi - z^* \rangle_{X_0^*, X_0} \geq 0, \text{ for all } \phi \in \mathcal{C} \\ \big(1 + \langle F_0'(\gamma^* z^*), z^* \rangle_{X_0^*, X_0} + \langle \mu, G_1'(\gamma^* z^*) z^* \rangle_{Y^*, Y}\big)(\gamma - \gamma^*) \geq 0, \text{ for all } \gamma \geq 0. \end{cases}$$

In terms of the variable y we have the optimality condition

(5.4)
$$\begin{cases} \langle F_0'(y^*) + G_1'(y^*)^* \mu, \phi - y^* \rangle_{X_0^*, X_0} \geq 0, \text{ for all } |\Lambda \phi|_{L^\infty} \leq \gamma^* \\ \big(\gamma^* + \langle F_0'(y^*), y^* \rangle_{X_0^*, X_0} + \langle \mu, G_1'(y^*) y^* \rangle_{Y^*, Y}\big)(\gamma - \gamma^*) \geq 0, \text{ for all } \gamma \geq 0. \end{cases}$$

If $\Lambda = Id$, $X_0 = L^\infty$ and $F_0'(y^*) + G_1'(y^*)^* \mu \in L^1(\Omega)$ then the variational inequality in (5.4) can be expressed as

$$(5.5) \quad \begin{cases} (F_0'(y^*) + G_1'(y^*)^* \mu)(x) = 0, & a.e. \text{ on } \{\, |y^*(x)| < \gamma^*\} \\ y^*(x) = -\gamma^* sgn(F_0'(y^*) + G_1'(y^*)^* \mu)(x)), & a.e. \text{ on } \{\, |y^*(x)| = \gamma^*\}. \end{cases}$$

where

$$sgn(s) \in \begin{cases} 1 \text{ for } s > 0 \\ -1 \text{ for } s < 0 \\ [-1, 1] \text{ for } s = 0. \end{cases}$$

6 A Class of State Constraint Problems

Without aiming for generality, we consider a non-smooth optimal control problem with distributed state constraints. Let A denote the generator of a semigroup $S(t)$ on a real Banach space X_0, let U denote the control space and $B \in \mathcal{L}(U, X_0)$ the control operator. We consider the linear control system on the fixed time horizon $[0, T]$

$$(6.1) \quad \begin{cases} \frac{d}{dt} x = Ax(t) + Bu(t), \text{ on } (0, T] \\ x(0) = x_0, \end{cases}$$

where $x_0 \in X_0$ and $u \in L^2(0, T; U)$. The solution to (6.1) is understood in the mild sense

$$x(t) = S(t)x_0 + \int_0^t S(t-s)Bu(s)\, ds.$$

The problem under consideration is

$$\text{(6.2)} \qquad \begin{cases} \min \int_0^T \big(\ell(x(t)) + h(u(t))\big)\, dt \\ \text{subject to (6.1) and } \int_0^T w(t) g(x(t))\, dt \le \delta, \end{cases}$$

where $\delta > 0$ and $w \in L^\infty(0,T)$ is a weight function. Here $\ell \in C(X_0, \mathbb{R})$, $h \in C(U, \mathbb{R})$, $g \in C(X_0, \mathbb{R})$ are assumed to be convex. It is further assumed that $x \to \int_0^T \ell(x(t))\, dt$, $x \to \int_0^T g(x(t))\, dt$, and $u \to \int_0^T h(u(t))\, dt$ are elements of $C(L^2(0,T;X_0), \mathbb{R})$ and $C(L^2(0,T;U), \mathbb{R})$ respectively. If moreover $g(0) = 0$, then the regular point condition will be satisfied.

Let (x^*, u^*) denote a solution to (6.2). To derive a necessary optimality condition we use the framework of Section 3 with

$$y = (x,u) \in X = L^2(0,T;X_0) \times L^2(0,T;U),\ Y = L^2(0,T;X_0),\ Z = \mathbb{R},\ K = \mathbb{R}^-,$$

and

$$\begin{aligned} F(y) &= \int_0^T \big(l(x(t)) + h(u(t)) \big)\, dt \\ G_1(y)(t) &= S(t)x_0 + \int_0^t S(t-s)Bu(s)\, ds - x(t) \\ G_2(y) &= \int_0^T w(t) g(x(t))\, dt - \delta. \end{aligned}$$

Since we assumed that $g(0) = 0$ and since $x \to \int_0^T w(t) g(x(t))\, dt$ is continuous on $L^2(0,T;X_0)$ the regular point condition can be argued. Hence by Theorem 3.3 there exists a Lagrange multiplier $(\lambda, \mu) \in L^2(0,T;X_0^*) \times \mathbb{R}^+$ associated to the solution (x^*, u^*). Before we show the optimality condition let us note that

$$G_1'(y^*)(h,v)(t) = \int_0^t S(t-s)Bv(s)\, ds - h(t)$$

and

$$\int_0^T \langle\, G_1'(y^*)(h,v)(t), \lambda(t) \,\rangle\, dt = -\int_0^T \langle\, \lambda(t), h(t) \,\rangle\, dt + \int_0^T (B^* p(t), v(t))\, dt,$$

where

$$\text{(6.3)} \qquad p(t) = \int_t^T S^*(s-t)\lambda(s)\, ds.$$

Thus by Theorem 3.3 we have

(6.4)

$$\int_0^T [\,\ell(x(t)) - \ell(x^*(t)) + \mu w(t)\,(g(x(t)) - g(x^*(t))) - \langle \lambda(t), x(t) - x^*(t)\,\rangle\,]\,dt$$

$$+ \int_0^T [\,h(v(t)) - h(u^*(t)) + (B^*p(t), v(t) - u^*(t))\,]\,dt \geq 0,$$

for all $x \in L^2(0,T;X_0)$ and $v \in L^2(0,T;U)$. Setting $v = u^*$ we find for a.e. $t \in (0,T)$, and all $u \in U$, $x \in X_0$

$$\ell(x) - \ell(x(t^*)) + \mu w(t)(g(x) - g(x^*(t))) - \langle \lambda(t), x - x^*(t)\,\rangle \geq\, 0.$$

Therefore $\lambda(t) \in \partial\big(\ell(x^*(t)) + \mu w(t)\, g(x^*(t))\big) \subset \partial\big(\ell(x^*(t))\big) + \mu\, w(t) \partial\big(g(x^*(t))\big)$. Hence there exist

$$q(t) \in \partial \ell(x^*(t)), \quad z(t) \in \partial g(x^*(t)),$$

with $q \in L^2(0,T;X_0^*)$, $z \in L^2(0,T;X_0^*)$, such that

$$\lambda(t) = q(t) + \mu w(t)\, z(t)$$

and

$$\int_t^T S^*(s-t)\lambda(s)\,ds = \int_t^T S^*(s-t)(q(s) + \mu w(s)\, z(s))\,ds.$$

By (6.3) therefore

$$p(t) = \int_t^T S^*(s-t)(q(s) + \mu w(s)\, z(s))\,ds, \tag{6.5}$$

which has the differential form

$$-\frac{d}{dt}p = A^*p(t) + q(t) + \mu w(t)\, z(t), \quad p(T) = 0.$$

If we let $x = x^*$, then (6.4) implies that

$$\int_0^T [\,h(v(t)) - h(u^*(t)) + (B^*p(t), v(t) - u^*(t))\,]\,dt \geq 0$$

for all $v \in L^2(0,T;U)$, and therefore

$$u^*(t) = \mathrm{argmin}_{v\in U}\big(\,h(v) + (B^*p(t), v)\,\big).$$

Summarizing we have the following result.

Theorem 6.1. *Suppose that* (x^*, u^*) *is an optimal solution to* (6.2). *Then there exists* $\mu \geq 0$ *and* $q(t) \in \partial \ell(x^*(t))$, $z(t) \in \partial G_2(x^*(t))$, *such that*

$$\mu \left(\int_0^T w(t) g(x^*(t))\, dt - \delta\right) = 0, \text{ and}$$
$$u^*(t) = \mathrm{argmin}_{v \in U}\big(h(v) + (B(x(t))^* p(t), v)\big),$$

where $p \in C(0,T; X_0^*)$ *satisfies the adjoint equation* (6.5).

7 Appendix

Consider

$$\mathcal{H}(t) = f(x(t)) + \delta |u(t)| + \big(b(x(t), u(t)), p(t)\big)_{\mathbb{R}^n}$$

along an optimal trajectory x^* and control u^* satisfying (4.1) with associated adjoint state p satisfying (4.7). For arbitrary $t \in (0,1)$ and h such that $t + h \in (0,t)$ we find, setting $f^0(x^*(t), u^*(t)) = f(x^*(t)) + \delta u^*(t)$,

$$\mathcal{H}(t+h) - \mathcal{H}(t) = \mathcal{H}(x^*(t), p(t), u^*(t+h)) - \mathcal{H}(x^*(t), p(t), u^*(t))$$
$$+ f^0(x^*(t+h), u^*(t+h)) - f^0(x^*(t), u^*(t+h))$$
$$+ (p(t+h) - p(t)) \cdot \tau^* b(x^*(t), u^*(t+h))$$
$$+ p(t+h) \cdot \big(\tau^* b(x^*(t+h), u^*(t+h)) - \tau^* b(x^*(t), u^*(t+h))\big)$$

Adding and subtracting $h\tau^* b(x^*(t+h), u^*(t+h)) \cdot \frac{d}{dt} p(t+h)$ on the right hand side, using $\mathcal{H}(x^*(t), p(t), u^*(t+h)) \geq \mathcal{H}(x^*(t), p(t), u^*(t))$ and the primary and adjoint equations we find

$$\mathcal{H}(t+h)-\mathcal{H}(t)$$

$$\geq f^0(x^*(t+h),u^*(t+h))-f^0(x^*(t),u^*(t+h))-hf^0_x(x^*(t+h),u^*(t+h))\tfrac{dx^*}{dt}(t+h)$$

$$+(p(t+h)-p(t))\cdot\tau^*b(x^*(t+h),u^*(t+h))-h\frac{dp}{dt}(t+h)\tau^*b(x^*(t+h),u^*(t+h))$$

$$+p(t+h)\cdot\tau^*\big(b(x^*(t+h),u^*(t+h))-b(x^*(t),u^*(t+h))$$

$$-hb_x(x^*(t+h),u^*(t+h))\frac{dx^*}{dt}(t+h)\big)$$

$$+(p(t+h)-p(t))\cdot\big((\tau^*b(x^*(t),u^*(t+h))-\tau^*b(x^*(t+h),u^*(t+h))\big).$$

Similarly,

$$\mathcal{H}(t)-\mathcal{H}(t+h)=\mathcal{H}(x^*(t+h),p(t+h),u^*(t+h))-\mathcal{H}(x^*(t+h),p(t+h),u^*(t))$$

$$+f^0(x^*(t+h),u^*(t))-f^0(x^*(t),u^*(t))$$

$$+(p(t+h)-p(t))\cdot\tau^*b(x^*(t+h),u^*(t))+p(t)\cdot(\tau^*b(x^*(t+h),u^*(t))-\tau^*b(x^*(t),u^*(t)))$$

$$\leq f^0(x^*(t+h),u^*(t))-f^0(x^*(t),u^*(t))-hf^0_x(x^*(t),u^*(t))\frac{dx^*}{dt}(t)$$

$$+(p(t+h)-p(t))\cdot\tau^*b(x^*(t),u^*(t))-h\frac{dp}{dt}(t)\cdot\tau^*b(x^*(t),u^*(t))$$

$$+p(t)\cdot(\tau^*b(x^*(t+h),u^*(t))-\tau^*b(x^*(t),u^*(t)))-hp(t)\cdot\tau^*b_x(x^*(t),u^*(t))\frac{dx^*}{dt}(t)$$

$$+(p(t+h)-p(t))\cdot\big(\tau^*b(x^*(t+h),u^*(t))-\tau^*b(x^*(t),u^*(t))\big).$$

Dividing these expressions by h and taking the limit $h\to 0$ we arrive at

$$\lim_{h\to 0}\frac{\mathcal{H}(t+h)-\mathcal{H}(t)}{h}=0.$$

Here we use that

$$b(x^*(t+h),u^*(t+h)) - b(x^*(t),u^*(t+h)) - b_x(x^*(t+h),u^*(t+h))\tfrac{d}{dt}x^*(t+h)$$

$$= \int_0^1 \big(b_x(x^*((1-s)t+s(t+h)))\tfrac{d}{dt}x^*((1-s)t+s(t+h))$$

$$-b_x(x^*(t+h),u^*(t+h))\tfrac{d}{ds}x^*(t+h)\big)\,ds \to 0, \text{ for } h\to 0,$$

and similarly for the remaining terms.

References

[C] E. Casas: Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations, SIAM. J. Control and Opti., 35(1997), 1297-1327.

[CY] E. Casas and J. Yong: Maximum principle for state-constrained optimal control problems governed by quasilinear elliptic equations, Differential and Integral Equations, 8(1995), 1-18.

[Cl] F.H. Clark: *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.

[CJK] C. Clason, B. Jin and K. Kunisch: A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration, to appear in SIAM Journal on Imaging Sciences.

[CK] C. Clason and K. Kunisch: A duality-based approach to elliptic control problems in non-reflexive Banach spaces, ESAIM: COCV, to appear.

[E] I. Ekeland: Nonconvex minimization problems, Bull. Am. Math. Soc. (N.S.), 1, (1979), 443-474.

[ET] I. Ekeland and R. Temam: *Convex Analysis and Variational Problems*, North Holland, Amsterdam, 1976.

[IK1] K. Ito and K.Kunisch: On the Lagrange Multiplier Approach to Variational Problems and Applications, SIAM, Philadelphia, 2008.

[LY] X. Li and J. Yong: Necessary conditions for optimal control of distributed parameter systems, SIAM. J. Control and Opti., 29(1991), 895-908.

[M] R.H. Martin: *Nonlinear Operators and Differential Equations in Banach Spaces*, Wiley&Sons, New York, 1976.

[MaZo] H. Maurer and J. Zowe: First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems, Mathematical Programming 16(1979), 98–110.

[WW] G. Wachsmuth and D. Wachsmuth: Convergence and Regularization Results for Optimal Control Problems with Sparsity Functional to appear in: ESAIM: Control, Optimisation and Calculus of Variations (COCV)

[Y] K. Yosida: *Functional Analysis*, Springer, Berlin, 1974.

[ZoKu] J. Zowe and S. Kurcyusz, Regularity and stability for the mathematical programming problem in Banach spaces, Appl. Math. Optim. 5(1979), 49–62.