SIAM J. CONTROL OPTIM.
Vol. 49, No. 5, pp. 2133–2154

KARUSH–KUHN–TUCKER CONDITIONS FOR NONSMOOTH MATHEMATICAL PROGRAMMING PROBLEMS IN FUNCTION SPACES*

KAZUFUMI ITO† AND KARL KUNISCH‡

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 nonsmooth cost.

Key words. KKT conditions, nonsmooth optimization, necessary optimality

AMS subject classifications. 49K21, 90C30, 90C46, 90C90

DOI. 10.1137/100817061

1. Introduction.

In this paper we discuss nonsmooth mathematical programming problems, which in part rely on convexity in part on regularity assumptions; more precisely we consider

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

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 of our research lies in 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 nonlinear equations are simpler to realize numerically than differential inclusions. If all operations in (P) are smooth, then the Maurer–Zowe–Kurcyusz [MaZo, ZoKu, IK1] conditions provide the Lagrangian framework that we are looking for. In the case that nondifferentiable 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 therein); the latter arise, for example, in the context of minimal effort optimal control problems.

*Received by the editors December 6, 2010; accepted for publication (in revised form) July 15, 2011; published electronically October 20, 2011.
http://www.siam.org/journals/sicon/49-5/81706.html

†Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 (kito@math.ncsu.edu). This author's research was partially supported by Army Research Office grants 56349-MA and 49308-MA and Air Force Research Office grant FA9550-09-1-0520.

‡Institut für Mathematik und wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, A-8010 Graz, Austria (karl.kunisch@uni-graz.at). This author's research was supported in part by the Fonds zur Förderung der wissenschaftlichen Forschung under SFB 32, "Mathematical Optimization and Applications in the Biomedical Sciences."

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 asymptotic expressions resulting from the variational principle. The approach follows that the multiplier rule can be of nonqualified form; i.e., there appears also a multiplier associated with the cost F. Special attention must be paid to guarantee that at least one of the multipliers is nontrivial and, in particular, that the multiplier associated with F is nontrivial. The results of this paper can also be compared to the abstract multiplier result obtained in [BC], which is obtained in the framework of Clark's differential calculus. The assumptions in [BC] use that F and G_1 are locally Lipschitzian and that G_1 has finite-dimensional range, and that the inequality constraint is of the form $G_2 \in K$, where K is a convex set with nonempty interior in a Banach space and G_2 is differentiable. To obtain a qualified form of the multiplier rule, it is assumed that G_1 is differentiable as well, and that a Slater-type condition holds, or alternatively, that no equality constraint is present and a family of properly perturbed optimization problems admits a solution. The study of variational problems with cost-functionals consisting of smooth and nondifferentiable, but convex, parts has a long history. We mention, for example, the monographs [GLT, Glo, IK1], where problems arising in the context of variational inequalities of the first and second kinds are investigated, and refer the reader to, e.g., [K] and the references therein for numerical methods.

We next briefly outline the contents of the paper. In section 2 the case where 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 nontriviality of the multipliers, regularity conditions are required. 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 to 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 assumed to hold throughout:

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

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*

$$\tag{2.1} \begin{cases} \lambda_0 \left(F_0'(y^*)(y - y^*) + F_1(y) - F_1(y^*) \right) + (\mu_1, G_1'(y^*)(y - y^*))_{\mathbb{R}^m} \\ \qquad\qquad + (\mu_2, G_2(y) - G_2(y^*))_{\mathbb{R}^p} \geq 0 \text{ for all } y \in \mathcal{C} \cap domF_1, \\ \mu_2 \geq 0, \quad G_2(y^*) \leq 0, \quad (\mu_2, G_2(y^*))_{\mathbb{R}^p} = 0. \end{cases}$$

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 defines, in a natural way, 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, p. 266]), there exists a $y^\epsilon \in \mathcal{C}$ such that

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

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

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

We further define

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

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

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

For the following estimate we use

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

where $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}$. Then from (2.4)

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

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

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

where $\|\cdot\|$ denotes the operator norm in $\mathcal{L}(X, \mathbb{R}^m)$. 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 $t \in (0, \bar{t}(\epsilon))$

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

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

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

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

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

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 $v \to |\max(0,v)|^2_{\mathbb{R}^p}$

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

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

$$
\begin{aligned}
&-\sqrt{\epsilon}\, d(y_t, y^\epsilon) \\
&\leq \frac{1}{J_\epsilon(y^\epsilon)+J_\epsilon(y_t)} \left[\tilde{\alpha}^{\epsilon,t}\big(tF_0'(y^\epsilon)(\hat{y}-y^\epsilon)+t(F_1(\hat{y})-F_1(y^\epsilon))\big)\right. \\
&\qquad \left. + 2\,(\tilde{\mu}_1^\epsilon, G_1'(y^\epsilon)(y_t-y^\epsilon))_{\mathbb{R}^m} + 2\,t\,\big(\tilde{\mu}_2^{\epsilon,t}, G_2(\hat{y})-G_2(y^\epsilon)\big)_{\mathbb{R}^p}\right] + o\,(|t|)
\end{aligned} \tag{2.9}
$$

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

$$
\alpha^{\epsilon,t} = \frac{\tilde{\alpha}^{\epsilon,t}}{J_\epsilon(y_t)+J_\epsilon(y^\epsilon)}, \quad \mu^{\epsilon,t} = \frac{2\,\tilde{\mu}^{\epsilon,t}}{J_\epsilon(y_t)+J_\epsilon(y^\epsilon)}, \quad \text{where } \tilde{\mu}^{\epsilon,t} = (\tilde{\mu}_1^\epsilon, \tilde{\mu}_2^{\epsilon,t}),\ \tilde{\mu}_2^{\epsilon,t} \geq 0.
$$

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

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

and

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

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

$$
\begin{aligned}
-\sqrt{\epsilon} d(\hat{y}, y^\epsilon) &\leq \lambda_{0,\epsilon}\,(F_0'(y^\epsilon)(\hat{y}-y^\epsilon)+F_1(\hat{y})-F_1(y^\epsilon)) \\
&\quad + (\mu_1^\epsilon, G_1'(y^\epsilon)(\hat{y}-y^\epsilon))_{\mathbb{R}^m} + (\mu_2^\epsilon, G_2(\hat{y})-G_2(y^\epsilon))_{\mathbb{R}^p}.
\end{aligned}
$$

Since $\{(\lambda_{0,\epsilon}, \mu^\epsilon) : \epsilon \in (0,1)\}$ is bounded in $\mathbb{R}\times\mathbb{R}^m\times\mathbb{R}^p$ as a consequence of (2.10), there exist $(\lambda_0, \mu) \in \mathbb{R}\times\mathbb{R}^m\times\mathbb{R}^p$ 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

$$
\begin{aligned}
0 &\leq \lambda_0\,\big(F_0'(y^*)(\hat{y}-y^*)+F_1(\hat{y})-F_1(y^*)\big) \\
&\quad + (\mu_1, G_1'(y^*)(\hat{y}-y^*))_{\mathbb{R}^m} + (\mu_2, G_2(\hat{y})-G_2(y^*))_{\mathbb{R}^p}
\end{aligned}
$$

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 real Banach spaces, and assume that Y^* and Z^* are strictly convex. Recall, for example, that $C^*(\Omega)$ is strictly convex [D]. 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

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

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, p. 50]. The convex subdifferential 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 $\hat{z} \notin K$. The assumption that Z^* is strictly convex is utilized to guarantee that $\partial d_K(\hat{z})$ is a singleton for any $\hat{z} \notin K$; see [LY]. Similarly, strict convexity of Y^* guarantees that the duality mapping from Y to Y^* is single valued.

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, L], where $G_2 : X \to Z$ is called convex with respect to the ordering introduced by the cone $K \subset Z$ if

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

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

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

then

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

Thus, if (3.2) holds, then d-convexity in the sense of Definition 3.1 and convexity as in (3.1) coincide. In the case of $Z = L^p(\Omega)$ or $Z = C(\Omega)$ and $K = \{z : z \leq 0 \text{ a.e.}\}$, 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 assumed to hold:

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

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 infinite-dimensional case, since the norm is only weakly lower semicontinuous, 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 and that the regular point condition*

$$0 \in \text{int}\,\{G_2(y) - K,\ y \in \mathcal{C} \cap domF_1\} \tag{3.4}$$

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

$$\begin{cases} \lambda_0\left(F_0'(y^*)(y-y^*) + F_1(y) - F_1(y^*)\right) + \left(\mu_1, G_1'(y^*)(y-y^*)\right)_{\mathbb{R}^m} \\ \qquad\qquad + \langle \mu_2, G_2(y) - G_2(y^*)\rangle_{Z^*,Z} \geq 0 \text{ for all } y \in \mathcal{C} \cap domF_1, \\ \langle \mu_2, z\rangle_{Z^*,Z} \leq 0 \text{ for all } z \in K,\ G_2(y^*) \leq 0,\ \langle \mu_2, G_2(y^*)\rangle_{Z^*,Z} = 0. \end{cases} \tag{3.5}$$

If $Y = \mathbb{R}^m$, *then is* $(\lambda_0, \mu_1, \mu_2)$ *nontrivial.*

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 case that $G_2(y^\varepsilon) \in K$ for all ε sufficiently small 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

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

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, p. 44] and [Y]), we have

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

For each fixed $\epsilon > 0$ consider $\{\tilde{\mu}_1^{\epsilon,t}\}_{t>0} = \{F(G_1(y_t))\}_{t>0}$ in Y^* as $t \to 0^+$. Since Y^* is strictly convex, the duality mapping is demicontinuous (see, e.g., [IK, p. 1]), and hence $F(G_1(y_t)) \to F(G_1(y^\epsilon))$ weakly* in Y^* as $t \to 0^+$. We set $\tilde{\mu}_1^\epsilon = F(G_1(y^\epsilon))$. For $t \to 0^*$ we have the estimate

$$\begin{aligned} &\langle \tilde{\mu}_1^{\epsilon,t}, G_1(y_t) - G_1(y^\epsilon)\rangle - t\langle \tilde{\mu}_1^\varepsilon, G_1'(y^\epsilon)(\hat{y} - y^\epsilon)\rangle \\ &= \langle \tilde{\mu}_1^{\epsilon,t}, G_1(y_t) - G_1(y^\epsilon) - tG_1'(y^\epsilon)(\hat{y} - y^\epsilon)\rangle\ +\ t\,\langle \tilde{\mu}_1^{\epsilon,t} - \tilde{\mu}_1^\epsilon, G_1'(y^\epsilon)(\hat{y} - y^\epsilon)\rangle = o(|t|), \end{aligned}$$

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

$$|G_1(y_t)|_Y^2 - |G_1(y^\epsilon)|_Y^2 \leq\ 2\,t\,\langle \tilde{\mu}_1^\epsilon, G'(y^\epsilon)(\hat{y} - y^\epsilon)\rangle_{Y^*,Y} + o(|t|), \tag{3.7}$$

where $\tilde{\mu}_1^\epsilon \in \partial\varphi(G_1(y^\epsilon)) = F(G_1(y^\epsilon))$.

For the third term we obtain

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

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

$$d_K(G_2(y_t))^2 - d_K(G_2(y^\varepsilon))^2 \le t\,\langle \tilde{\mu}_2^{\epsilon,t}, G_2(\hat{y}) - G_2(y^\epsilon)\rangle_{Z^*,Z}. \tag{3.9}$$

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

$$\begin{aligned}
&-\sqrt{\epsilon}\, d(y_t, y^\epsilon)\\
&\le \tfrac{1}{J_\epsilon(y^\epsilon)+J_\epsilon(y_t)}\left[\tilde{\alpha}^{\epsilon,t}\big(tF_0'(y^\epsilon)(\hat{y}-y^\epsilon) + t(F_1(\hat{y}) - F_1(y^\epsilon))\big)\right.\\
&\quad + 2\,t\,\langle \tilde{\mu}_1^\epsilon, G_1'(y^\epsilon)(\hat{y}-y^\epsilon)\rangle_{Y^*,Y} + t\,\langle \tilde{\mu}_2^{\epsilon,t}, G_2(\hat{y}) - G_2(y^\epsilon)\rangle_{Z^*,Z}] + o\,(|t|)
\end{aligned} \tag{3.10}$$

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

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

We next consider a sequence $\xi_{t_n} \in \partial d_K(G_2(y_{t_n}))$ as $t_n \to 0^+$. Since $G_2(y^\varepsilon)$ is not in K by assumption and since K is assumed to be closed, it follows that $G_2(y_{t_n}) \notin K$ for all n sufficiently large. As a consequence, $|\xi_{t_n}|_{Z^*} = 1$ for all such n. The following argument is analogous to that for establishing demicontinuity of the duality mapping; see, e.g., [IK, p. 2]. Since closed balls in Z^* are w^*-compact, there exists a w^*-accumulation point ξ^ϵ of ξ_{t_n}. The set $\{G_2(y_{t_n})\}_{n=1}^\infty \cup \{G_2(y^\epsilon\}$ is a separable subspace of Z. We momentarily restrict ourselves to this closed separable subspace of Z. Since for separable spaces the w^*-topology on w^*-compact subsets of Z^* is metrizable, we deduce the existence of a subsequence such that $w^* - \lim \xi_{t_{n_k}} = w^* - \lim \partial d_K(G_2(y_{t_{n_k}})) = \partial d_K(G_2(y^\epsilon)) = \xi^\varepsilon$.

We further 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 $|\xi^\varepsilon|_{Z^*} = 1$ is a consequence of $\xi^\varepsilon \in \partial d_K(G_2(y^\varepsilon))$, 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, $\lambda_{0,\epsilon} \ge 0$. Dividing (3.10) by t and letting $t \to 0^+$, we obtain

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

Since $\{(\lambda_{0,\epsilon}, \mu_1^\epsilon, \mu_2^\epsilon) : \epsilon \in (0,1)\}$ is bounded in $\mathbb{R} \times Y^* \times Z^*$, there exist $(\lambda_0, \mu_1, \mu_2) \in \mathbb{R} \times Y^* \times Z^*$ 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

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

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

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

Since K is a convex cone, we have $y + G_2(y^*) \in K$ for any $y \in K$. This implies that $\langle \mu_2, y\rangle \leq 0$ for all $y \in K$ and in particular $\langle \mu_2, G_2(y^*)\rangle \leq 0$. Setting $y = 0$ in (3.15) we have $\langle \mu_2, G_2(y^*)\rangle \geq 0$, 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. Then $|\xi_t|_{Z^*} \leq 1$ for all $t \geq 0$. 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 nontriviality 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 $\langle \xi^\varepsilon, y - G_2(y^\varepsilon)\rangle_{Z^*,Z} \leq 0$ for all $y \in K$, and hence

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

From (3.13) we deduce that

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

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

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

From (3.4) it follows that there exists $0 \neq z_0 \in Z$ 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

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

By (3.16) we obtain

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

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 $0 \in \text{int}\{G_2(y) - G_2(y^*) : y \in \mathcal{C} \cap domF_1\}$, 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

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

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

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

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)

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

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

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

which allows us 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*

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

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

Proof. In view of the proof of the previous theorem, we need only 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

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

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

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

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

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

Hence $\mu_1 = 0$ is impossible. ∎

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

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

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)

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

Using (3.17)

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

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

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

This implies that

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

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

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

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

$$\begin{aligned}&\langle \mu_1, \hat\mu_1\rangle_{Y^*,Y} + \langle \mu_2, \hat\mu_2\rangle_{Z^*,Z}\\ &\quad \ge \langle \mu_1, G_1'(y^*)(y - y^*)\,\rangle_{Y^*,Y} + \langle\, \mu_2, G_2(y) - G_2(y^*)\,\rangle_{Z^*,Z} \ge 0\end{aligned}$$

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 $\lambda^0 = 1$. ▯

4. L^1-minimum norm control.

Consider the time-optimal, L^1-minimum norm problem

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

where $\delta > 0$, $x_0 \in \mathbb{R}^n$, $|\cdot|_{\mathbb{R}^k}$ denotes the Euclidean norm in $\mathbb{R}^k$, 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.

In the context of sparse controls the pointwise norm constraints allow us to avoid controls in measure space. In fact, an L^1-cost without constraints on the controls does not guarantee existence of a minimizer in L^1; rather, we would need to resort to a formulation with measure-valued controls. We refer the reader to [CK] for the treatment of this topic in the context of optimal control of elliptic equations with sparsity constraints, and also to [St, WW] for a treatment of L^1-controls with L^2-regularization.

Let us assume for a moment that we have a single input system, i.e., $k = 1$. Then, as we shall see from the optimality system (4.10) below, except for a set of critical values of the adjoint, the optimal control assumes the values $0, \pm\mu$, and thus it is of "bang-bang-off" type.

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

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

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

We impose that the regular point condition

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

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) with

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

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 exists a Lagrange multiplier $(\lambda_0, \mu) \in \mathbb{R}^+ \times \mathbb{R}^m$ such that

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

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

$$\begin{aligned} &\tau^*\,(F_1(u) - F_1(u^*)) + (\tau - \tau^*)\, F_1(u^*) \\ &\quad + \big((F_0(y^*))_u + \mu^T G_u(y^*)\big)(u - u^*) + \big((F_0(y^*))_\tau + \mu^T G_\tau(y^*)\big)(\tau - \tau^*) \geq 0 \end{aligned} \tag{4.5}$$

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

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

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

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

where (h, ξ) satisfies

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

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

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

Then

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

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

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

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

$$\begin{aligned} &(\tau - \tau^*) \int_0^1 \left(f(x^*(t)) + \delta|u^*(t)| + (b(x^*(t), u^*(t)), p(t))_{\mathbb{R}^n}\right) dt \\ &+ \tau^* \int_0^1 \left((b_u(x^*(t), u^*(t))^t p(t), u(t) - u^*(t))_{\mathbb{R}^k} + \delta|u(t)| - \delta|u^*(t)| \right) dt \geq 0. \end{aligned} \tag{4.8}$$

We first set $\tau = \tau^*$ and choose $t_0 \in (0,1)$ as a common Lebesgue point of $b_u(x^*, u^*)^t p$ and u^*. For a $\rho > 0$ sufficiently small so that the ball $B(t_0, \rho) \in (0,1)$ we set

$$u_\rho(t) = \begin{cases} u \in B(t_0, \rho), \\ u^*(t) \in (0,1) \setminus B(t_0, \rho), \end{cases}$$

where $u \in \mathbb{R}^m$ satisfies $|u| \leq \gamma$. Passing to the limit in

$$\begin{aligned} \lim_{\rho \to 0^+} \frac{1}{|B(t_0, \rho)|} \int_{B(t_0,\rho)} &\left((b_u(x^*(t), u^*(t))^t p(t), u_\rho(t) - u^*(t))_{\mathbb{R}^k} \right. \\ &\left. + \delta|u_\rho(t)| - \delta|u^*(t)| \right) dt \geq 0, \end{aligned}$$

and observing that the set of common Lebesgue points for $b_u(x^*, u^*)^t p$ and u^* has measure 1 we obtain for almost every $t \in (0,1)$

$$(b_u(x^*(t), u^*(t))^t p(t), u - u^*(t))_{\mathbb{R}^k} + \delta|u| - \delta|u^*(t)| \geq 0 \text{ for all } |u| \leq \gamma. \tag{4.9}$$

This implies that

$$u^*(t) \in \begin{cases} 0 & \text{if } |b_u(x^*(t), u^*(t))^T p(t)| < \delta, \\ -[0,\gamma]\, \frac{b_u(x^*(t),u^*(t))^T p(t)}{|b_u(x^*(t),u^*(t))^T p(t)|} & \text{if } |b_u(x^*(t), u^*(t))^T p(t)| = \delta, \\ -\gamma\, \frac{b_u(x^*(t),u^*(t))^T p(t)}{|b_u(x^*(t),u^*(t))^T p(t)|} & \text{if } |b_u(x^*(t), u^*(t))^T p(t)| > \delta. \end{cases} \tag{4.10}$$

In fact, we note that (4.9) is the necessary optimality condition for

$$\min_{|v|_{\mathbb{R}^k} \leq \gamma} (b_u(x^*(t), u^*(t))^T p(t), v)_{\mathbb{R}^k} + \delta |v|_{\mathbb{R}^k}. \tag{4.11}$$

Setting $q(t) = b_u(x^*(t), u^*(t))^T p(t)$, the minimum is obtained for some $v = -\alpha \frac{q(t)}{|q(t)|}$ with $\alpha \in [0, \gamma]$, and we can equivalently express (4.11) as

$$\min_{\alpha \in [0,\gamma]} \alpha(-|q(t)|_{\mathbb{R}^k} + \delta).$$

For the solution we obtain

$$\alpha^*(t) \in \begin{cases} 0 & \text{if } |q(t)| < \delta, \\ [0,\gamma] & \text{if } |q(t)| = \delta, \\ \gamma & \text{if } |q(t)| > \delta. \end{cases}$$

This justifies (4.10).

We reconsider (4.8) and set $u = u^*$ in (4.8) next. We obtain

$$\int_0^1 \big(f(x^*(t)) + \delta|u^*(t)| + (b(x^*(t), u^*(t)), p(t))_{\mathbb{R}^n}\big)\, dt = 0. \tag{4.12}$$

The Hamiltonian associated with (4.2), given by

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

is constant along the optimal trajectory x^* and control u^*. For a proof of this fact we refer to the appendix. Together with (4.12) this implies that

$$\mathcal{H}(x^*(t), u^*(t), p(t)) = f(x^*(t)) + \delta|u^*(t)| + (b(x^*(t), u^*(t)), p(t))_{\mathbb{R}^n} = 0 \text{ on } [0,1]. \tag{4.13}$$

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) *hold. Then, there exists a Lagrange multiplier* $\mu \in \mathbb{R}^m$ *such that the optimality conditions* (4.10), (4.14) *hold, where* $p \in H^1((0,1); R^n)$ *satisfies the adjoint equation* (4.7).

Remark 4.1. If in (4.1) the L^1 term is replaced by $\frac{\delta}{2} \int_0^\tau |u(t)|^2\, dt$, then the optimal control is found to satisfy for almost every $t \in (0, T)$

$$u^*(t) = P_{B_\gamma}\left(-\frac{1}{\delta} b_u(x^*(t), u^*(t))^T p(t)\right), \tag{4.14}$$

where P_{B_γ} denotes the projection onto $\{v \in \mathbb{R}^k : |v|_{\mathbb{R}^k} \leq \gamma\}$. Comparing the effect of the L^1- and L^2-cost terms for this constrained optimal control problem, we note that for $|b_u(x^*(t), u^*(t))^T p(t)| > \delta$ the controls coincide, while for $|b_u(x^*(t), u^*(t))^T p(t)| < \delta$ the L^1-cost, differently from the L^2-cost, is sparse.

Remark 4.2. In the case when g is the identity, the regular point condition holds if the linearized nonautonomous control system

$$(4.15) \qquad \begin{cases} \frac{d}{dt} h(t) = \tau^*(b_x(x^*(t), u^*(t))h(t) + b_u(x^*(t), u^*(t))(u(t) - u^*(t))), \\ h(0) = 0 \ \text{ with } u \in U_{ad} \end{cases}$$

is controllable with constraints on the controls, in the sense that $0 \in \operatorname{int} R(1)$, where $R(1) = \{h(1; u) : u \in U_{ad}\}$ is the reachable set at $t = 1$. 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 $0 \in \operatorname{int} R(1)$; see, e.g., [LM, section 2.2].

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 expressed as

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

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 $\mathcal{C} = \{z \in X_0 : |\Lambda z|_{L^\infty} \leq 1\}$. 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

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

is assumed 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

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

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

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

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-constrained problems. Without aiming for generality, we consider a nonsmooth 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 let $B \in \mathcal{L}(U, X_0)$ be the control operator. We consider the linear control system on the fixed time horizon $[0, T]$

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

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 as

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

The problem under consideration is

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

where $\vec{\delta} \in \mathbb{R}^p$, and $w \in L^\infty(0, T; \mathbb{R}^{p\times m})$ is a weight function. Here $\ell \in C^1(X_0, \mathbb{R})$, $h \in C(U, \mathbb{R})$, $g \in C(X_0, \mathbb{R}^m)$ are assumed to be convex. It is further assumed that $x \to \int_0^T \ell(x(t))\,dt$, $x \to \int_0^T w(x)\,g(x(t))\,dt$, and $u \to \int_0^T h(u(t))\,dt$ are elements of $C(L^2(0, T; X_0), \mathbb{R})$, $C(L^2(0, T; X_0), \mathbb{R}^m)$, and $C(L^2(0, T; U), \mathbb{R})$, respectively.

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

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

and

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

LEMMA 6.1. *If there exists* $v \in L^2(0,T;U)$ *with* $G_2\big(x^* + \int_0^{\cdot} S(\cdot - s)Bv(s)\,ds\big) < \vec{\delta}$, *then the regular point condition* (3.20) *holds.*

Proof. We need to argue that

$$0 \in \operatorname{int}\left\{ \begin{pmatrix} \mathcal{S}v - (x - x^*) \\ \mathcal{G}(x) - \vec{\delta} - \vec{k} \end{pmatrix} : (x,\, v) \in X,\, \vec{k} \in (\mathbb{R}^-)^p \right\}, \tag{6.3}$$

where the right-hand side is considered as a subset of $L^2(0,T;X_0) \times \mathbb{R}^p$, and

$$(\mathcal{S}v)(t) = \int_0^t S(t-s)Bv(s)\,ds, \quad \mathcal{G}(x) = \int_0^T w(t)g(x(t))\,dt.$$

To verify (6.3) we need to solve the following equation for arbitrary $(z_1, \vec{z}_2) \in L^2(0,T;X_2) \times (\mathbb{R})^p$ sufficiently small:

$$\begin{cases} \mathcal{S}v - (x - x^*) = z_1, \\ \mathcal{G}(x) - \vec{\delta} - \vec{k} = \vec{z}_2, \end{cases} \tag{6.4}$$

where $\vec{k} \in (\mathbb{R}^-)^p$. This is equivalent to solving

$$\mathcal{G}(x^* - z_1 + \mathcal{S}v) - \vec{\delta} - \vec{k} = \vec{z}_2.$$

Let v be chosen as in the statement of the assumption of the lemma. Then

$$\mathcal{G}(x^* + \mathcal{S}v) < \vec{\delta}.$$

Hence there exists $\rho > 0$ such that for $(z_1, \vec{z}_2)$ with $|(z_1, \vec{z}_2)|_{L^2(0,T;X_0)\times\mathbb{R}^p} < \rho$, we have

$$\mathcal{G}(x^* - z_1 + \mathcal{S}v) - \vec{\delta} - \vec{z}_2 \le \vec{0}.$$

Thus $\vec{k} = \mathcal{G}(x^* - z_1 + \mathcal{S}v) - \vec{\delta} - \vec{z}_2$ provides the solution to (6.4). ∎

By Theorem 3.3 and the above lemma there exists a Lagrange multiplier $(\lambda, \vec{\mu}) \in L^2(0,T;X_0^*) \times (\mathbb{R}^+)^p$ associated with 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

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

where

$$p(t) = \int_t^T S^*(s-t)\lambda(s)\,ds. \tag{6.5}$$

Thus by Theorem 3.3 we have

$$\begin{aligned} &\int_0^T [\,\ell(x(t)) - \ell(x^*(t)) + \vec{\mu}^T w(t)\,(g(x(t)) - g(x^*(t))) - \langle \lambda(t), x(t) - x^*(t)\,\rangle\,]\,dt \\ &\quad + \int_0^T [\,h(v(t)) - h(u^*(t)) + (B^*p(t), v(t) - u^*(t))_U\,]\,dt \ge 0 \end{aligned} \tag{6.6}$$

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 $u \in U$, $x \in X_0$,

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

Therefore $\lambda(t) \in \partial\big(\ell(x^*(t)) + \vec{\mu}^T w(t)\, g(x^*(t))\big) = \partial\big(\ell(x^*(t))\big) + \vec{\mu}^T\, w(t)\partial\big(g(x^*(t))\big)$; see, e.g., [ET, p. 26]. Hence there exist

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

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

$$\lambda(t) = q(t) + \vec{\mu}^T w(t)\, z(t)$$

and

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

By (6.5) therefore

$$p(t) = \int_t^T S^*(s-t)(q(s) + \vec{\mu}^T w(s)\, z(s))\,ds, \tag{6.7}$$

which has the differential form

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

If we let $x = x^*$, then (6.6) implies that

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

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

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

Summarizing we have the following result.

Theorem 6.1. *Suppose that $y^* = (x^*, u^*)$ is a local solution to* (6.2) *and that $v \to G_2'(x^*)(\int_0^\cdot S(\cdot - s)Bv(s)\,ds)$ is nontrivial. Then there exist $\vec{\mu} \in (\mathbb{R}^+)^p$, $q(t) \in \partial\ell(x^*(t))$, and $z(t) \in \partial G_2(x^*(t))$ such that*

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

where $p \in C(0,T;X_0^*)$ satisfies the adjoint equation (6.7).

7. Appendix.

Consider

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

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

$$\begin{aligned}
&\mathcal{H}(x^*(t+h),p(t+h),u^*(t+h)) - \mathcal{H}(x^*(t),p(t),u^*(t))\\
&= \mathcal{H}(x^*(t),p(t),u^*(t+h)) - \mathcal{H}(x^*(t),p(t),u^*(t))\\
&\quad + f(x^*(t+h),u^*(t+h)) - f(x^*(t),u^*(t+h))\\
&\quad + (p(t+h)-p(t))\cdot b(x^*(t),u^*(t+h))\\
&\quad + p(t+h)\cdot\big(b(x^*(t+h),u^*(t+h)) - b(x^*(t),u^*(t+h))\big).
\end{aligned}$$

Adding and subtracting $hb(x^*(t+h),u^*(t+h))\cdot\frac{d}{dt}p(t+h)$ on the right-hand side, and 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

$$\begin{aligned}
&\mathcal{H}(x^*(t+h),p(t+h),u^*(t+h)) - \mathcal{H}(x^*(t),p(t),u^*(t))\\
&\geq f(x^*(t+h),u^*(t+h)) - f(x^*(t),u^*(t+h)) - hf_x(x^*(t+h),u^*(t+h))\frac{dx^*}{dt}(t+h)\\
&\quad + (p(t+h)-p(t))\cdot b(x^*(t+h),u^*(t+h)) - h\frac{dp}{dt}(t+h)b(x^*(t+h),u^*(t+h))\\
&\quad + p(t+h)\cdot\Bigg(b(x^*(t+h),u^*(t+h)) - b(x^*(t),u^*(t+h))\\
&\quad - hb_x(x^*(t+h),u^*(t+h))\frac{dx^*}{dt}(t+h)\Bigg)\\
&\quad + (p(t+h)-p(t))\cdot\big((b(x^*(t),u^*(t+h)) - b(x^*(t+h),u^*(t+h))\big).
\end{aligned}$$

Here we used that

$$\begin{aligned}
&b(x^*(t+h),u^*(t+h)) - b(x^*(t),u^*(t+h)) - b_x(x^*(t+h),u^*(t+h))\frac{d}{dt}x^*(t+h)\\
&\quad = \int_0^1 \Bigg(b_x(x^*((1-s)t+s(t+h)))\frac{d}{dt}x^*((1-s)t+s(t+h))\\
&\qquad\qquad - b_x(x^*(t+h),u^*(t+h))\frac{d}{ds}x^*(t+h)\Bigg)\,ds \to 0 \text{ for } h\to 0,
\end{aligned}$$

and similarly for the remaining terms.

Similarly,

$$\begin{aligned}
&\mathcal{H}(x^*(t+h),p(t+h),u^*(t+h)) - \mathcal{H}(x^*(t),p(t),u^*(t))\\
&\qquad = \mathcal{H}(x^*(t+h),p(t+h),u^*(t+h)) - \mathcal{H}(x^*(t+h),p(t+h),u^*(t))\\
&\qquad\quad + f(x^*(t+h),u^*(t)) - f(x^*(t),u^*(t))\\
&\quad + (p(t+h)-p(t))\cdot b(x^*(t+h),u^*(t)) + p(t)\cdot(b(x^*(t+h),u^*(t)) - b(x^*(t),u^*(t)))
\end{aligned}$$

$$
\begin{aligned}
&\le f(x^*(t+h),u^*(t)) - f(x^*(t),u^*(t)) - hf_x(x^*(t),u^*(t))\frac{dx^*}{dt}(t)\\
&+ (p(t+h)-p(t))\cdot b(x^*(t),u^*(t)) - h\frac{dp}{dt}(t)\cdot b(x^*(t),u^*(t))\\
&+ p(t)\cdot(b(x^*(t+h),u^*(t)) - b(x^*(t),u^*(t))) - hp(t)\cdot b_x(x^*(t),u^*(t))\frac{dx^*}{dt}(t)\\
&+ (p(t+h)-p(t))\cdot\big(b(x^*(t+h),u^*(t)) - b(x^*(t),u^*(t))\big).
\end{aligned}
$$

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

$$
\lim_{h\to 0} \frac{\mathcal{H}(x^*(t+h),p(t+h),u^*(t+h)) - \mathcal{H}(x^*(t),p(t),u^*(t))}{h} = 0.
$$

REFERENCES

[BC] J. F. Bonnas and E. Casas, *Controle de systemes elliptiques semi-linear comportant des constraints distribuees sur l'etat*, in Nonlinear Differential Equations and Their Applications, College de France Seminar VIII, H. Brezis and J. L. Lions, eds., Pitman, Boston, 1988, pp. 69–86.

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

[CY] E. Casas and J. Yong, *Maximum principle for state-constrained optimal control problems governed by quasilinear elliptic equations*, Differential Integral Equations, 8 (1995), pp. 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*, SIAM J. Imaging Sci., 3 (2010), pp. 199–231.

[CK] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in nonreflexive Banach spaces*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 243–266.

[D] M. M. Day, *Strict convexity and smoothness of normed spaces*, Trans. AMS, 78 (1955), pp. 516–528.

[E] I. Ekeland, *Nonconvex minimization problems*, Bull. Amer. Math. Soc. (N.S.), 1 (1979), pp. 443–474.

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

[GLT] R. Glowinski, J. L. Lions, and R. Tremoliers, *Numerical Analysis of Variational Inequalities*, North–Holland, Amsterdam, 1981.

[Glo] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer-Verlag, Berlin, 1984.

[IK] K. Ito and F. Kappel, *Evolution Equations and Approximations*, World Scientific, River Edge, NJ, 2002.

[IK1] K. Ito and K. Kunisch, *Lagrange Multiplier Approach to Variational Problems and Applications*, Adv. Des. Control 15, SIAM, Philadelphia, 2008.

[K] K. C. Kiwiel, *A method for minimizing the sum of a convex function and a continuously differentiable function*, J. Optim. Theory Appl., 48 (1986), pp. 437–449.

[LM] E. B. Lee and L. Markus, *Foundations of Optimal Control Theory*, John Wiley & Sons, New York, 1967.

[LY] X. Li and J. Yong, *Necessary conditions for optimal control of distributed parameter systems*, SIAM J. Control Optim., 29 (1991), pp. 895–908.

[L] D. G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons, New York, 1969.

[M] R. H. Martin, *Nonlinear Operators and Differential Equations in Banach Spaces*, John 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*, Math. Programming, 16 (1979), pp. 98–110.

[St] G. Stadler, *Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2009), pp. 159–181.

[WW] G. Wachsmuth and D. Wachsmuth, *Convergence and regularization results for optimal control problems with sparsity functional*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 858–886.

[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), pp. 49–62.