OPTIMAL CONTROL PROBLEMS WITH MIXED CONSTRAINTS

FRANCIS CLARKE* AND M. R. DE PINHO†

Abstract. We develop necessary conditions of broad applicability for optimal control problems in which the state and control are subject to mixed constraints. We unify, subsume and significantly extend most of the results on this subject, notably in the three special cases that comprise the bulk of the literature: calculus of variations, differential-algebraic systems, and mixed constraints specified by equalities and inequalities. Our approach also provides a new and unified calibrated formulation of the appropriate constraint qualifications, and shows how to extend them to nonsmooth data. Other features include a very weak hypothesis concerning the type of local minimum, nonrestrictive hypotheses on the data, and stronger conclusions, notably as regards the maximum (or Weierstrass) condition. The necessary conditions are stratified, in the sense that they are asserted on precisely the domain upon which the hypotheses (and the optimality) are assumed to hold. This leads to local, intermediate, and global versions of the necessary conditions, according to how the hypotheses are formulated.

Key words. optimal control, necessary conditions, mixed constraints, nonsmooth analysis

AMS subject classifications. 49K15, 49K21

1. Introduction.

We study in this article an optimal control problem with standard cost and dynamics, but in which the state x and control u are subject to joint, or *mixed constraints* through the condition $(x(t), u(t)) \in S(t)$. The presence of such constraints has long been known to constitute a challenge as regards the derivation of appropriate necessary conditions of maximum principle type. Problems with mixed constraints have been studied systematically by Hestenes [23], Dubovickiĭ and Milyutin [21], Gamkrelidze [22] and Neustadt [28] among many others, and remain an active subject: see [1, 2, 6, 7, 12, 14, 15, 17, 18, 19, 20, 26, 27, 29, 30, 31, 33, 34].

In [7], Clarke presents a synthesis of what has been called the *nonsmooth analysis approach* to necessary conditions in optimal control. The results constitute, from several points of view, the current state of the art for standard optimal control problems. Although the issue of mixed constraints is broached, it is not completely developed. The purpose of this article is to do so. The principal result, Theorem 2.1 below, is a set of necessary conditions obtained under a geometric hypothesis called the *bounded slope condition.* A functional form of the theorem is also derived (Theorem 4.3).

It turns out that these theorems unify and significantly extend most of the existing results, notably in the three special cases that comprise the bulk of the literature: calculus of variations (Section 5), differential-algebraic systems (Section 6), and mixed constraints specified by equalities and inequalities (Section 7). Our approach also provides a new and unified *calibrated* formulation of the appropriate constraint qualifications (Section 4), and shows how to extend them to nonsmooth data. Other features include a very weak hypothesis concerning the type of local minimum, nonrestrictive hypotheses on the data (on convexity, regularity, and boundedness), and stronger conclusions (notably as regards the maximum – or Weierstrass – condition), even for problems with smooth data. The necessary conditions are *stratified*, in the sense that they are asserted to exactly the same extent that the hypotheses (and the

*Institut universitaire de France and Université de Lyon, Institut Camille Jordan, CNRS UMR 5208, 69622 Villeurbanne, France (clarke@math.univ-lyon1.fr).

† ISR and DEEC, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal (mrpinho@fe.up.pt).

optimality) are assumed to hold, as specified by the presence of a *radius function.* This leads to local, intermediate, and global versions of the results, according to how the hypotheses are framed.

We proceed to review for the reader's convenience some basic definitions from nonsmooth analysis. Given a nonempty closed subset S of $\mathbb{R}^n$ and a point x in S, we say that $\zeta \in \mathbb{R}^n$ is a *proximal normal* (vector) to S at x if there exists $\sigma = \sigma(x, \zeta) \geq 0$ such that

$$\langle \zeta, x' - x \rangle \;\leq\; \sigma |x' - x|^2 \quad \forall\, x' \in S.$$

This is the *proximal normal inequality.* The set of such ζ, which is a convex cone containing 0, is denoted $N_S^P(x)$, and is referred to as the proximal normal cone. It can be viewed as the basic building block of the proximal theory of nonsmooth analysis (which applies to smooth Banach spaces, such as $\mathbb{R}^n$). Given a lower semicontinuous function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and a point x at which f is finite, we say that ζ is a *proximal subgradient* of f at x if there exists $\sigma = \sigma(x) \geq 0$ and a neighborhood V_x of x such that

$$f(x') - f(x) + \sigma |x' - x|^2 \;\geq\; \langle \zeta, x' - x \rangle \quad \forall\, x' \in V_x.$$

The set of such ζ, which may be empty, is denoted $\partial_P f(x)$ and referred to as the proximal subdifferential. The *limiting normal cone* $N_S^L(x)$ to S at x is obtained by applying a sequential closure operation to N_S^P:

$$N_S^L(x) := \big\{ \lim \zeta_i : \zeta_i \in N_S^P(x_i),\ x_i \to x,\ x_i \in S \big\}.$$

A similar procedure defines the *limiting subdifferential*:

$$\partial_L f(x) := \big\{ \lim \zeta_i : \zeta_i \in \partial_P f(x_i),\ x_i \to x,\ f(x_i) \to f(x) \big\}.$$

In an arbitrary Banach space, one can develop nonsmooth calculus via the theory of *generalized gradients.* In the case of a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}$, the generalized gradient $\partial_C f(x)$ coincides with $\operatorname{co} \partial_L f(x)$; further, the associated normal cone $N_S^C(x)$ to a set S at a point x coincides with $\overline{\operatorname{co}}\, N_S^L(x)$.

If the set S is convex, the three normal cones defined above coincide with the familiar normal cone of convex analysis, in which ζ is normal to S at x iff $\langle \zeta, x' - x \rangle \leq 0 \;\; \forall\, x' \in S$. If S is a smooth manifold, or manifold with boundary, then $N_S^L(x)$ and $N_S^C(x)$ coincide with the classical normal space, or half-space. If f is convex, then $\partial_P f(x)$, $\partial_L f(x)$ and $\partial_C f(x)$ all coincide with the subdifferential of convex analysis. If the function f is strictly differentiable at x (in particular, if f is C^1 near x), then $\partial_L f(x) = \partial_C f(x) = \{f'(x)\}$.

Because of these facts, the theorem statements in this article can all be understood without reference to nonsmooth analysis, by simply assuming the data of the underlying problems to be smooth, in which case the generalized derivatives and normals in the statements coincide with the usual ones. We stress, however, that nonsmooth calculus plays an essential role in the proofs; this would still be the case even if only problems with smooth data were considered, and even in that case the results of this article furnish a new state of the art. We refer to [10] for a detailed exposition of nonsmooth calculus in the notation used here.

2. A geometric theorem on mixed constraints. We are given an interval $[a,b]$ in $\mathbb{R}$ and a subset S of $[a,b]\times\mathbb{R}^n\times\mathbb{R}^m$. We write

$$S(t) := \big\{(x,u) : (t,x,u) \in S\big\}, \quad S(t,x) := \big\{u : (t,x,u) \in S\big\}.$$

Also given are a subset E of $\mathbb{R}^n\times\mathbb{R}^n$ together with functions

$$f : [a,b]\times\mathbb{R}^n\times\mathbb{R}^m \to \mathbb{R}^n,\ \ \Lambda : [a,b]\times\mathbb{R}^n\times\mathbb{R}^m \to \mathbb{R},\ \ \ell : \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}.$$

We consider the following problem (P) of optimal control :

$$(P) \quad \left\{ \begin{array}{rcll} \multicolumn{4}{l}{\text{Minimize } J(x,u) \ := \ \ell(x(a),x(b)) + \int_a^b \Lambda_t(x(t),u(t))\,dt} \\ \multicolumn{4}{l}{\text{subject to}} \\ x'(t) & = & f_t(x(t),u(t)) & \text{a.e.} \quad t \in [a,b] \\ (x(t),u(t)) & \in & S(t) & \text{a.e.} \quad t \in [a,b] \\ (x(a),x(b)) & \in & E. & \end{array} \right.$$

Notice that the t-dependence of Λ and f is reflected by means of a subscript. This will be convenient for notational reasons, and should cause no confusion, since no partial derivatives with respect to t are ever taken. The *basic hypotheses* on the problem data, in force throughout, are the following: f,Λ are $\mathcal{L}\times\mathcal{B}$ measurable[1]; S is $\mathcal{L}\times\mathcal{B}$ measurable; E is closed; ℓ is locally Lipschitz.

It is understood that this problem involves measurable control functions $u(t)$ and absolutely continuous functions $x(t)$ (arcs). Such a pair (or process) (x,u) is said to be *admissible* for (P) if the constraints are satisfied and $J(x,u)$ is well-defined and finite. The main theorem features hypotheses directly related to a given pair (x_*,u_*) that is admissible for (P). Let $R : [a,b] \to (0,+\infty]$ be a given measurable *radius function*, and $\epsilon > 0$. We say that (x_*,u_*) is a *local minimum of radius* R for (P) provided that for every pair (x,u) admissible for (P) which also satisfies

$$\big|x(t) - x_*(t)\big| \le \epsilon,\ \big|u(t) - u_*(t)\big| \le R(t) \text{ a.e.}, \qquad \int_a^b \big|x'(t) - x_*'(t)\big|\,dt \le \epsilon,$$

we have $J(x,u) \ge J(x_*,u_*)$. This resembles a so-called $W^{1,1}$ local minimum, which is known to be a weaker hypothesis than the classical strong local minimum. But it is even weaker than that notion, because of the additional restriction stemming from the radius function. (This restriction vanishes, of course, if R is identically $+\infty$, which is allowed.)

The two main hypotheses of the theorem are conditioned by the radius R; they concern *Lipschitz behavior* of f and Λ with respect to (x,u) and a certain *bounded slope condition* bearing upon the sets $S(t)$. Both of these are imposed only at points $(x,u) \in S(t)$ that are near (x_*,u_*), as determined by ϵ and R in the following definition:

$$S_*^{\epsilon,R}(t) := \big\{(x,u) \in S(t) : |x - x_*(t)| \le \epsilon,\ |u - u_*(t)| \le R(t)\big\}.$$

In generic terms, we shall say that any given function $\phi_t(x,u)$ satisfies $\mathrm{L}_*^{\epsilon,R}$ if the following holds:

$\mathbf{L}_*^{\epsilon,\mathbf{R}}$: There exist measurable real-valued functions k_x^ϕ, k_u^ϕ such that, for almost every t in $[a,b]$, for every (x_i,u_i) in a neighborhood of $S_*^{\epsilon,R}(t)$ $(i=1,2)$, we have

$$\big|\phi_t(x_1,u_1) - \phi_t(x_2,u_2)\big| \ \le \ k_x^\phi(t)\big|x_1 - x_2\big| + k_u^\phi(t)\big|u_1 - u_2\big|.$$

[1]This hypothesis, familiar in control theory, refers to measurability relative to the σ-field generated by the products of Lebesgue measurable subsets in $\mathbb{R}$ and Borel measurable subsets in $\mathbb{R}^n\times\mathbb{R}^m$.

Thus, when we require later that f satisfy $\mathrm{L}_*^{\epsilon,R}$, for example, the associated Lipschitz parameters are denoted by k_x^f, k_u^f. Concerning the mixed constraint, the hypothesis is the following:

$\mathbf{BS}_*^{\epsilon,\mathbf{R}}$: There exists a measurable real-valued function k_S such that, for almost every t, the following bounded slope condition holds:

$$(x,u) \in S_*^{\epsilon,R}(t),\ (\alpha,\beta) \in N_{S(t)}^P(x,u) \implies |\alpha| \ \le\ k_S(t)|\beta|.$$

We assume that $S(t)$ is locally closed at each point $(x,u) \in S_*^{\epsilon,R}(t)$, so that the normal cone appearing above is well-defined.

The following theorem asserts necessary conditions under optimality and regularity hypotheses which are imposed only for a radius R, and whose conclusions hold to the same extent; this situation is referred to in [7] as *stratified*.

THEOREM 2.1. *Let (x_*, u_*) be a local minimum of radius R for (P), where* $\mathrm{BS}_*^{\epsilon,R}$ *holds, where f and Λ satisfy* $\mathrm{L}_*^{\epsilon,R}$*, where the functions*

$$k_x^f,\ \ k_x^\Lambda,\ \ k_S\left[k_u^f + k_u^\Lambda\right]$$

are summable, and where, for some $\eta > 0$, we have $R(t) \ge \eta k_S(t)$ a.e.

1. *Then there exist an arc p and a number λ_0 in $\{0,1\}$ satisfying the* **nontriviality condition**

$$(\lambda_0, p(t)) \ne 0 \quad \forall\, t \in [a,b],$$

the **transversality condition**

$$(p(a), -p(b)) \ \in\ \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N_E^L(x_*(a), x_*(b)),$$

the **Euler adjoint inclusion** *for almost every t:*

$$(-p'(t), 0) \ \in\ \partial_C \big\{ \langle p(t), f_t\rangle - \lambda_0 \Lambda_t \big\}(x_*(t), u_*(t)) - N_{S(t)}^C(x_*(t), u_*(t)),$$

as well as the **Weierstrass condition of radius R** *for almost every t:*

$$\begin{aligned}&(x_*(t), u) \in S(t),\ |u - u_*(t)| \le R(t) \implies \\ &\langle p(t), f_t(x_*(t), u)\rangle - \lambda_0 \Lambda_t(x_*(t), u) \ \le\ \langle p(t), f_t(x_*(t), u_*(t))\rangle - \lambda_0 \Lambda_t(x_*(t), u_*(t)).\end{aligned}$$

2. *Further, if the hypotheses hold for a sequence of radius functions R_i going to $+\infty$ (with all parameters ϵ, k_x^f, k_u^f, k_x^Λ, k_u^Λ, k_S, η possibly depending on i), in the sense that* $\liminf_{i\to\infty} R_i(t) = +\infty$ a.e., *then the conclusions above hold for an arc p which satisfies the* **global Weierstrass condition** *for almost every t:*

$$\begin{aligned}&(x_*(t), u) \in S(t) \implies \\ &\langle p(t), f_t(x_*(t), u)\rangle - \lambda_0 \Lambda_t(x_*(t), u) \ \le\ \langle p(t), f_t(x_*(t), u_*(t))\rangle - \lambda_0 \Lambda_t(x_*(t), u_*(t)).\end{aligned}$$

Examples show that the hypothesis $R(t) \ge \eta k_S(t)$ a.e. is needed in the Theorem (see [7, p. 47]), as well as the summability conditions. We remark that the bounded slope condition excludes unilateral state constraints; that is, constraints of the type $x(t) \in X(t)$. In such a case, we have

$$S(t) \ =\ \{(x,u) : x \in X(t)\}\,,$$

a set admitting proximal normals of the form $(\alpha, 0)$ with $\alpha \neq 0$, which makes the bounded slope condition impossible. It is well-known that in the presence of such constraints, necessary conditions of the type given above fail, and that their appropriate extensions involve measures and adjoint arcs p that are discontinuous (see for example [35]).

The theorem is discussed further in Section 8, in the context of comparing our results with those in the literature. Its proof is given in Section 9.

3. The unmixed case, and a hybrid theorem. In this section we consider first the implications of Theorem 2.1 for the case in which $S(t) = \{(x, u) : u \in U(t)\}$. Here the control constraints are unilateral or unmixed (but see the hybrid result below), and the problem (P) of Section 2 reduces to a classical optimal control problem incorporating the dynamics and control constraints

$$x'(t) = f_t(x(t), u(t)),\ u(t) \in U(t) \text{ a.e.}$$

Let (α, β) belong to $N^P_{S(t)}(x, u)$. Then, by definition of proximal normal, for some constant σ, the function

$$(y, w) \mapsto -\langle \alpha, y \rangle - \langle \beta, w \rangle + \sigma \left\{ |y - x|^2 + |w - u|^2 \right\}$$

has a local minimum relative to $(y, w) \in \mathbb{R}^n \times U(t)$ at $(y, w) = (x, u)$. It follows that $\alpha = 0$, so that the bounded slope condition $\mathrm{BS}^{\epsilon,R}_*$ of Section 2 is automatically satisfied, with $k_S = 0$, for any radius R. We assume that f, Λ satisfy the Lipschitz condition $\mathrm{L}^{\epsilon,R}_*$, and that k^f_x and k^Λ_x are summable. Then, if U is measurable and closed-valued, Theorem 2.1 applies, and we obtain:

Theorem 3.1. *If (x_*, u_*) is a local minimum of radius R, then there exist p and λ_0 satisfying all the conclusions of Theorem 2.1, where the adjoint inclusion is expressible in the form*

$$(-p'(t), 0) \in \partial_C \left\{ \langle p(t), f_t \rangle - \lambda_0 \Lambda_t \right\} (x_*(t), u_*(t)) - \{0\} \times N^C_{U(t)}(u_*(t)),$$

and where the Weierstrass condition of radius R holds for almost every t:

$$\begin{aligned} & u \in U(t),\ |u - u_*(t)| \leq R(t) \Longrightarrow \\ & \langle p(t), f_t(x_*(t), u) \rangle - \lambda_0 \Lambda_t(x_*(t), u) \ \leq\ \langle p(t), f_t(x_*(t), u_*(t)) \rangle - \lambda_0 \Lambda_t(x_*(t), u_*(t)). \end{aligned}$$

We remark that when f, Λ are locally Lipschitz, the above theorem applies for any finite constant radius R if one makes the classically familiar hypotheses that u_* is bounded and that the local minimum is a strong one ($W^{1,1}$ would also suffice). In that case, by letting $R \to +\infty$, we may assert the *global* Weierstrass condition. If we further specialize by requiring the data to be smooth in x, we recover the conclusions of the usual Pontryagin maximum principle [25, 32, 35].

Notice that Theorem 3.1 features a *coupled* adjoint equation involving derivatives with respect to (x, u) jointly, and referred to as the Euler adjoint inclusion. As regards necessary conditions that incorporate this form of the adjoint equation, Theorem 3.1 goes beyond previous results in various ways: since $k_S = 0$, we can allow *any* (arbitrarily small) positive time-dependent radius function, and still assert the Euler adjoint equation, as well as a local Weierstrass condition (absent in previous analyses). In addition, the functions k^f_u, k^Λ_u need not be summable.

As in the classical maximum principle, the well-known *nonsmooth maximum principle* [3, 4] features an adjoint inclusion that is *decoupled* from the control variable:

$$-p'(t) \in \partial_C \big\{ \langle p(t), f_t(\cdot, u_*(t)) \rangle - \lambda_0 \Lambda_t(\cdot, u_*(t)) \big\}(x_*(t)),$$

the generalized gradient being taken solely with respect to x. The relative merits of the two different forms of the adjoint inclusion are discussed in detail in [8, 13, 16, 24]. The Euler adjoint inclusion of Theorem 3.1 implies the decoupled one under some circumstances, notably when f, Λ are smooth in x. But they are distinct conclusions in general. There is a further distinction, however: Lipschitz behavior with respect to u was not a required hypothesis in obtaining the earlier decoupled result, nor was U necessarily closed-valued. For this reason, Theorem 2.1 does not directly subsume the original nonsmooth maximum principle. To remedy this, we prove a hybrid theorem that has more complicated hypotheses than Theorem 2.1, but which has the merit of fully subsuming the nonsmooth maximum principle for the case of unmixed control constraints (as well as Theorem 2.1).

A hybrid theorem. We consider the following hybrid problem (HP), in which the control u is partitioned into two components: $u = (v, w)$.

$$(HP) \quad \left\{ \begin{array}{rcll} \multicolumn{4}{l}{\text{Minimize } J(x,u) = J(x,v,w) \ := \ \ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), u(t))\, dt} \\ \multicolumn{4}{l}{\text{subject to}} \\ x'(t) & = & f_t(x(t), u(t)) & \text{a.e.} \quad t \in [a,b] \\ (x(t), v(t)) & \in & S(t, w(t)), \ w(t) \in W(t) & \text{a.e.} \quad t \in [a,b] \\ (x(a), x(b)) & \in & E. & \end{array} \right.$$

Observe that in (HP), only the w-component is subject to unilateral constraints as given by the multifunction W, while the mixed constraints involving v are dependent on w. The multifunctions $W(t)$ and $S(t, w)$ are taken to be $\mathcal{L} \times \mathcal{B}$ measurable.

We assume that $(x_*, u_*) = (x_*, v_*, w_*)$ is a local minimum of radius R for (HP): for every $(x, u) = (x, v, w)$ admissible for (HP) which also satisfies

$$\big| x(t) - x_*(t) \big| \le \epsilon, \ \big| (v(t), w(t)) - (v_*(t), w_*(t)) \big| \le R(t) \text{ a.e.}, \quad \int_a^b \big| x'(t) - x_*'(t) \big|\, dt \le \epsilon,$$

we have $J(x, u) \ge J(x_*, u_*)$. We turn now to the main hypotheses, which are asymmetrical in the v and w components for the reasons explained above; they impose no Lipschitz behavior with respect to w, nor continuity. We define

$$S_*^{\epsilon, R}(t, w) := \big\{ (x, v) \in S(t, w) : |x - x_*(t)| \le \epsilon, \ |v - v_*(t)| \le R(t) \big\}.$$

In generic terms, we shall say that any given function $\phi_t(x, v, w)$ satisfies $\mathrm{HL}_*^{\epsilon, R}$ if the following holds:

$\mathbf{HL}_*^{\epsilon,\mathbf{R}}$: There exist $\mathcal{L} \times \mathcal{B}$ measurable real-valued functions k_x^ϕ, k_v^ϕ such that, for almost every t in $[a, b]$, for every $w \in W(t) \cap B(w_*(t), R(t))$, for every (x_i, v_i) in a neighborhood of $S_*^{\epsilon, R}(t, w)$, $(i = 1, 2)$, we have

$$\big| \phi_t(x_1, v_1, w) - \phi_t(x_2, v_2, w) \big| \le k_x^\phi(t, w) \big| x_1 - x_2 \big| + k_v^\phi(t, w) \big| v_1 - v_2 \big|.$$

We assume that for each $w \in W(t) \cap B(w_*(t), R(t))$, the set $S(t, w)$ is locally closed at each point $(x, v) \in S_*^{\epsilon, R}(t, w)$. The bounded slope condition now becomes:

$\mathbf{HBS}_*^{\epsilon,\mathbf{R}}$: There exists a measurable real-valued function k_S such that, for almost every t, for every $w \in W(t) \cap B(w_*(t), R(t))$, the following bounded slope condition holds:

$$(x,v) \in S_*^{\epsilon,R}(t,w),\ (\alpha,\beta) \in N^P_{S(t,w)}(x,v) \implies |\alpha| \ \le\ k_S(t)|\beta|.$$

THEOREM 3.2. *Let (x_*, u_*) be a local minimum of radius R for (HP), where* $\mathrm{HBS}_*^{\epsilon,R}$ *holds, where f and Λ satisfy* $\mathrm{HL}_*^{\epsilon,R}$, *where the functions*

$$k_x^f(t,w_*(t)),\ \ k_x^\Lambda(t,w_*(t)),\ \ k_S(t)\big[k_v^f(t,w_*(t)) + k_v^\Lambda(t,w_*(t))\big]$$

are summable, and where, for some $\eta > 0$, we have $R(t) \ge \eta k_S(t)$ a.e.

1. *Then there exist an arc p and a number λ_0 in $\{0,1\}$ satisfying the* **nontriviality condition**

$$(\lambda_0, p(t)) \ne 0 \quad \forall\, t \in [a,b],$$

the **transversality condition**

$$(p(a), -p(b)) \ \in\ \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N^L_E(x_*(a), x_*(b)),$$

the **hybrid adjoint inclusion** *for almost every t:*

$$\begin{aligned}&(-p'(t), 0) \ \in \\ &\partial_C\big\{ \langle p(t), f_t(\cdot,\cdot,w_*(t))\rangle - \lambda_0 \Lambda_t(\cdot,\cdot,w_*(t))\big\}(x_*(t),v_*(t)) - N^C_{S(t,w_*(t))}(x_*(t),v_*(t)),\end{aligned}$$

as well as the **Weierstrass condition of radius R** *for almost every t:*

$$\begin{aligned}&w \in W(t),\ (x_*(t), v) \in S(t,w),\ |(v,w) - u_*(t)| \le R(t) \Longrightarrow \\ &\langle p(t), f_t(x_*(t),v,w)\rangle - \lambda_0\Lambda_t(x_*(t),v,w) \le \langle p(t), f_t(x_*(t),u_*(t))\rangle - \lambda_0\Lambda_t(x_*(t),u_*(t)).\end{aligned}$$

2. *Further, if the hypotheses hold for a sequence of radius functions R_i going to $+\infty$ (with all parameters ϵ, k_x^f, k_v^f, k_x^Λ, k_v^Λ, k_S, η possibly depending on i), in the sense that* $\liminf_{i\to\infty} R_i(t) = +\infty$ a.e., *then the conclusions above hold for an arc p which satisfies the* **global Weierstrass condition** *for almost every t:*

$$\begin{aligned}w \in W(t),\ (x_*(t), v) \in S(t,w) \Longrightarrow\ &\langle p(t), f_t(x_*(t),v,w)\rangle - \lambda_0\Lambda_t(x_*(t),v,w) \\ &\le \langle p(t), f_t(x_*(t),u_*(t))\rangle - \lambda_0\Lambda_t(x_*(t),u_*(t)).\end{aligned}$$

Notice that in the hybrid adjoint inclusion above, the generalized gradient is taken with respect to (x,v) jointly (but not w). When $W(t)$ is taken to be $\{0\}$ for every t (or equivalently, when w is absent), Theorem 3.2 reduces to Theorem 2.1. At the other extreme, when f and Λ are independent of v, the mixed constraint is irrelevant, and we may take $S(t,w)$ to be the whole space, and $k_S \equiv 0$. Then, for $R(t) \equiv +\infty$, the theorem reduces to the nonsmooth maximum principle. The reductions are exact with respect to the hypotheses as well as the conclusions. Theorem 3.2 is proved in Section 9.

4. Special structure of the constraint set. When the mixed constraint $(x,u) \in S(t)$ has special structure, it turns out to be possible in many cases to conveniently specify conditions in terms of that structure that imply the bounded slope condition needed in Theorem 2.1. In addition, such structure may give rise to a more explicit adjoint equation by providing an interpretation of the normal cone $N^C_{S(t)}$ via multipliers. We now develop some results along these lines, in a framework general enough to subsume certain more familiar ones to be visited later. We record the following fact from nonsmooth calculus:[2]

PROPOSITION 4.1. *Let Φ be a closed subset of $\mathbb{R}^N$, and let $\phi : \mathbb{R}^n \to \mathbb{R}^N$ be a function which is Lipschitz in a neighborhood of a point u_* satisfying $\phi(u_*) \in \Phi$. We set $S := \{u : \phi(u) \in \Phi\}$, and we suppose that*

$$\lambda \in N^L_\Phi(\phi(u_*)),\ 0 \in \partial_L \langle \lambda, \phi \rangle (u_*) \Longrightarrow \lambda = 0. \tag{4.1}$$

1. *If $\zeta \in N^L_S(u_*)$, there exists $\lambda \in N^L_\Phi(\phi(u_*))$ such that $\zeta \in \partial_L \langle \lambda, \phi \rangle (u_*)$.*
2. *If ϕ is strictly differentiable at u_*, and if $\zeta \in N^C_S(u_*)$, then there exists $\lambda \in N^C_\Phi(\phi(u_*))$ such that $\zeta = \nabla \langle \lambda, \phi \rangle (u_*)$.*

The hypothesis (4.1) is a variant of what is known in mathematical programming as a *constraint qualification* serving to rule out the abnormal case in the Lagrange multiplier rule (see for example Section 6.3 of [5]).

A calibrated constraint qualification. We now return to the problem (P), under the assumption throughout this section that, for a given function $\phi : [a,b] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^N$ and measurable mapping Φ from $[a,b]$ to the closed subsets of $\mathbb{R}^N$, the set $S(t)$ of Section 2 is described by

$$S(t) := \big\{(x,u) : \phi_t(x,u) \in \Phi(t)\big\}. \tag{4.2}$$

We postulate the same measurability and Lipschitz behavior for ϕ as we did for f, Λ; in particular, we assume that ϕ satisfies $\mathrm{L}^{\epsilon,R}_*$. Thus ϕ is locally Lipschitz in (x,u) at relevant points, but not necessarily in t. We are given as in Section 2 a process (x_*, u_*) admissible for the problem (P). The following *calibrated constraint qualification* relative to (x_*, u_*) will prove relevant:

$\mathbf{M}^{\epsilon,R}_*$ **:** There exists $M : [a,b] \to \mathbb{R}$ measurable such that, for almost every t,

$$(x,u) \in S^{\epsilon,R}_*(t),\ \lambda \in N^L_{\Phi(t)}(\phi_t(x,u)),\ (\alpha,\beta) \in \partial_L \langle \lambda, \phi_t \rangle (x,u) \Longrightarrow\ |\lambda| \le M(t)|\beta|.$$

PROPOSITION 4.2. *If $\mathrm{M}^{\epsilon,R}_*$ holds, then $\mathrm{BS}^{\epsilon,R}_*$ holds with $k_S(t) := M(t) k^\phi_x(t)$.*

Proof. Let t be such that the $\mathrm{M}^{\epsilon,R}_*$ property holds, as well as the Lipschitz condition of ϕ. Let

$$(x,u) \in S^{\epsilon,R}_*(t),\ (\alpha,\beta) \in N^P_{S(t)}(x,u).$$

By Prop. 4.1, there exists λ such that

$$\lambda \in N^L_{\Phi(t)}(\phi_t(x,u)),\ (\alpha,\beta) \in \partial_L \langle \lambda, \phi_t \rangle (x,u).$$

Then $|\lambda| \le M(t)|\beta|$ by $\mathrm{M}^{\epsilon,R}_*$. In addition, we have $|\alpha| \le |\lambda| k^\phi_x(t)$, since the function $y \mapsto \langle \lambda, \phi_t \rangle (y,v)$ is Lipschitz with constant $|\lambda| k^\phi_x(t)$ for (y,v) near (x,u). We deduce

$$|\alpha| \ \le\ M(t) k^\phi_x(t) |\beta|,$$

[2]This is derivable, for example, from the proximal chain rule [10, Theorem 9.1].

as required. □

As an immediate application of the results of this section, we derive the following consequence of Theorem 2.1 when the constraints are described as in (4.2).

THEOREM 4.3. *Let* (x_*, u_*) *be a local minimum of radius* R *for* (P), *where* $\mathrm{M}_*^{\epsilon,R}$ *holds, where* f, Λ, ϕ *satisfy* $\mathrm{L}_*^{\epsilon,R}$, *where the functions*

$$k_x^f,\ k_x^\Lambda,\ Mk_x^\phi\left[k_u^f + k_u^\Lambda\right]$$

are summable, and where, for some $\eta > 0$, *we have* $R(t) \geq \eta M(t) k_x^\phi(t)$ *a.e. Then there exist* p *and* λ_0 *satisfying all the conclusions of Theorem 2.1. Furthermore, if* ϕ_t *is strictly differentiable at* $(x_*(t), u_*(t))$ a.e., *then there exists a measurable function* $\lambda : [a, b] \to \mathbb{R}^N$ *satisfying*

$$\lambda(t) \in N^C_{\Phi(t)}(\phi_t(x_*(t), u_*(t))) \text{ a.e.}$$

such that the adjoint inclusion takes the explicit multiplier form

$$(-p'(t), 0) \in \partial_C \big\{ \langle p(t), f_t \rangle - \lambda_0 \Lambda_t - \langle \lambda(t), \phi_t \rangle \big\}(x_*(t), u_*(t)) \text{ a.e.}$$

The multiplier λ *satisfies*

$$|\lambda(t)| \leq M(t)\{|p(t)| k_u^f(t) + \lambda_0 k_u^\Lambda(t)\} \text{ a.e.}$$

In the case of a sequence of radius functions R_i *going to* $+\infty$, *this bound on* $|\lambda(t)|$ *holds for the data* M, k_u^f, k_u^Λ *corresponding to any one of the radius functions.*

Proof. Note that our hypotheses imply the local closedness of $S(t)$ at points in $S_*^{\epsilon,R}(t)$, as required in Theorem 2.1. According to Prop. 4.2, $\mathrm{BS}_*^{\epsilon,R}$ holds with $k_S(t) = M(t)k_x^\phi(t)$. In light of this, the first assertion of the Theorem is a direct consequence of Theorem 2.1. When ϕ_t is strictly differentiable, any element of $N^C_{S(t)}(x_*(t), u_*(t))$ is expressible in the form

$$\nabla \langle \lambda(t), \phi_t \rangle\, (x_*(t), u_*(t)),$$

in view of Prop. 4.1. In addition, we have by the sum rule

$$\partial_C\{\langle p(t), f_t\rangle - \lambda_0\Lambda_t - \langle \lambda(t), \phi_t\rangle\} = \partial_C\{\langle p(t), f_t\rangle - \lambda_0\Lambda_t\} - \nabla\langle \lambda(t), \phi_t\rangle\,.$$

There results the explicit adjoint inclusion of the Theorem. (That $\lambda(t)$ may be chosen measurably is an exercise in measurable selection theory; see for example [10, pp. 149-152].) It now follows from this adjoint inclusion that $\nabla\langle\lambda(t), \phi_t\rangle\,(x_*(t), u_*(t))$ is of the form (q, r), with $|r| \leq \{|p| k_u^f + \lambda_0 k_u^\Lambda\}$, where k_u^f and k_u^Λ correspond to any choice of radius function for which the hypotheses hold. This, together with $\mathrm{M}_*^{\epsilon,R}$ (for that same radius), yields the given estimate for $|\lambda(t)|$. □

We remark that the degree to which the explicit multiplier λ is summable is clearly identified in our approach: it belongs to the same L^p space (if any) as $M\{k_u^f + k_u^\Lambda\}$. Note that $\lambda(t) = 0$ a.e. at points t for which $\phi_t(x_*(t), u_*(t)) \in \operatorname{int}\Phi(t)$ (since the normal cone at such points reduces to $\{0\}$).

It turns out that various types of constraint qualifications encountered in the literature are subsumed by the calibrated condition $\mathrm{M}_*^{\epsilon,R}$ introduced above (which has the further merit of applying to nonsmooth data). We proceed in the rest of this section to make this explicit in some important special cases.

The first case invokes the *generalized Jacobian* $\partial_J \phi$ (see [5]), which reduces to the usual Jacobian matrix if ϕ is continuously differentiable.

PROPOSITION 4.4. *Let there exist $c > 0$ such that, for almost every t,*

$$(x,u) \in S_*^{\epsilon,R}(t),\ (A,B) \in \partial_J \phi_t(x,u) \implies \det BB^T \ \geq\ c.$$

Then $\mathrm{M}_*^{\epsilon,R}$ holds with a constant M.

Proof. The right-hand side above implies that BB^T is positive definite, so there exists $d > 0$ (independent of t) such that $\langle BB^T\lambda, \lambda\rangle \geq d|\lambda|^2 \ \ \forall\, \lambda \in \mathbb{R}^N$. Let t be such that the hypothesis holds, and let

$$(x,u) \in S_*^{\epsilon,R}(t),\ (\alpha,\beta) \in \partial_L \langle \lambda, \phi_t\rangle (x,u).$$

Then [5, 2.6.6] there exists $(A,B) \in \partial_J \phi_t(x,u)$ such that $\beta = B^T\lambda$. We deduce $|\beta|^2 \geq d|\lambda|^2$, which gives $\mathrm{M}_*^{\epsilon,R}$ with $M := d^{-1/2}$. □

The advantage of having $\mathrm{M}_*^{\epsilon,R}$ hold with a constant M lies in more easily confirming the summability requirements of Theorem 4.3. The next case also leads to this conclusion:

PROPOSITION 4.5. *Let ϕ be continuously differentiable on a neighborhood of S, and let $\Phi(t)$ be convex-valued. If for some positive r, c we have, for almost every t, for every $(x,u) \in S_*^{\epsilon,R}(t)$:*

$$B(0,r) \subset \big\{\phi_t(x,u) + \langle D_u\phi_t(x,u), u'\rangle : |u'| \leq c\big\} - \Phi(t),$$

then $\mathrm{M}_*^{\epsilon,R}$ holds for the constant $M = c/r$.

Proof. Let t be such that the hypothesis holds, and let

$$(x,u) \in S_*^{\epsilon,R}(t),\ 0 \neq \lambda \in N^L_{\Phi(t)}(\phi_t(x,u)),\ (\alpha,\beta) = \nabla\langle\lambda,\phi_t\rangle(x,u).$$

By hypothesis, there exists $\phi' \in \Phi$ and $u' \in B(0,c)$ such that $r\lambda/|\lambda| = \phi_t(x,u) + \langle D_u\phi_t(x,u), u'\rangle - \phi'$. Bearing in mind that $\langle \lambda, \phi' - \phi_t(x,u)\rangle \leq 0$ (normality in the convex sense), and taking inner products with λ, we discover

$$r|\lambda| \ \leq\ |\nabla_u\langle\lambda,\phi_t\rangle(x,u)\cdot u'| \ \leq\ c|\beta|,$$

which establishes $\mathrm{M}_*^{\epsilon,R}$ with $M := c/r$. □

We turn our attention next to (uncalibrated) constraint qualifications of a type often referred to as *Mangasarian-Fromowitz conditions.* We say that MFC holds at $(t,x,u) \in S$ provided that

$$\lambda \in N^L_{\Phi(t)}(\phi_t(x,u)),\ (\alpha,0) \in \partial_L\langle\lambda,\phi_t\rangle(x,u) \implies \lambda = 0. \tag{MFC}$$

The goal below is to identify two easily-recognized scenarios under which this simpler hypothesis implies $\mathrm{M}_*^{\epsilon,R}$. For this purpose we set

$$C_*^{\epsilon,R} := \mathrm{cl}\,\big\{(t,x,u) \in [a,b]\times\mathbb{R}^n\times\mathbb{R}^m : (x,u) \in S_*^{\epsilon,R}(t)\big\}.$$

Note that in defining this set, a specific value of $u_*(t)$ (as well as the t-dependent data R, ϕ, Φ) must be assigned for every t, not just almost everywhere.[3]

PROPOSITION 4.6. *Let Φ be autonomous and $C_*^{\epsilon,R}$ compact. Suppose that* MFC *holds at every $(t,x,u) \in C_*^{\epsilon,R}$. Then $\mathrm{M}_*^{\epsilon,R}$ holds for a constant M in either of the following cases:*

[3]The ensuing results are sharpest when, for every t, an *essential value* is picked, in the sense of [11]; we do not pursue this here, however.

1. *ϕ is autonomous and k_x^ϕ is constant.*

2. *In a neighborhood of $C_*^{\epsilon,R}$, $D_u\phi$ exists, and ϕ, $D_u\phi$ are continuous in (t,x,u).*

Proof. We proceed to prove the first case by contradiction. Suppose that for each i there exist $t_i \in [a,b]$, $(x_i,u_i) \in S_*^{\epsilon,R}(t_i)$, $\lambda_i \in N_\Phi^L(\phi(x_i,u_i))$, and $(\alpha_i,\beta_i) \in \partial_L\langle\lambda_i,\phi\rangle(x_i,u_i)$ such that $|\lambda_i| > i|\beta_i|$. By normalizing and taking subsequences, we may take $|\lambda_i| = 1$, and we may suppose $(t_i,x_i,u_i) \to (t,x,u) \in C_*^{\epsilon,R}$, $\lambda_i \to \lambda$, where λ is a unit vector in $N_\Phi^L(\phi(x,u))$. It follows that $\beta_i \to 0$. We also have $|\alpha_i| \le k_x^\phi$ [10, 1.7.3], so that we may assume $\alpha_i \to \alpha$. Then in the limit we derive $(\alpha,0) \in \partial_L\langle\lambda,\phi\rangle(x,u)$, which contradicts MFC at the point (t,x,u). The proof of the second case is similar. Suppose again that for each i there exist $(x_i,u_i) \in S_*^{\epsilon,R}(t_i)$, $\lambda_i \in N_\Phi^L(\phi_{t_i}(x_i,u_i))$, and $(\alpha_i,\beta_i) \in \partial_L\langle\lambda_i,\phi_{t_i}\rangle(x_i,u_i)$ such that $|\lambda_i| > i|\beta_i|$. Then $\beta_i = \nabla_u\langle\lambda_i,\phi_{t_i}\rangle(x_i,u_i)$. Once more, we may take $|\lambda_i| = 1$, and we may suppose $(t_i,x_i,u_i) \to (t,x,u) \in C_*^{\epsilon,R}$, $\lambda_i \to \lambda$, where λ is a unit vector in $N_\Phi^L(\phi_t(x,u))$. We have $\beta_i \to 0$, and $0 = \nabla_u\langle\lambda,\phi_t\rangle(x,u)$ in the limit, contradicting MFC at (t,x,u). □

It is natural to ask whether the constraint qualification could be imposed only *along* the optimal process. To this end, we introduce

DEFINITION 4.7. *We say that $(t,x_*(t),u)$ is an* admissible cluster point *of (x_*,u_*) if there exists a sequence $t_i \in [a,b]$ converging to t and corresponding points $(x_i,u_i) \in S(t_i)$ such that $\lim x_i = x_*(t)$ and $\lim u_i = \lim u_*(t_i) = u$. We say that M_*^0 holds if* MFC *holds at all admissible cluster points of (x_*,u_*).*

Note that if u_* is continuous, then $u = u_*(t)$ above, and, if all the data are continuous in t, the definition amounts to imposing MFC at $(t,x_*(t),u_*(t))$ for every t; that is, the Mangasarian-Fromowitz condition *along* the optimal process (x_*,u_*). When the t-behavior is discontinuous, however, the definition takes account of limit points in a way that can be shown to be essential for the necessary conditions to hold (see the example in Section 8). Note that in this setting we have as before assigned a value to $u_*(t)$ for *every* t, not just almost everywhere.

In the following, the constraint qualification is imposed only with reference to the optimal process, in the sense of Definition 4.7. Naturally, the conclusion is weaker: we obtain the Weierstrass condition only for a sufficiently small radius $\delta > 0$. This is new, however, relative to the existing literature.

THEOREM 4.8. *Let (x_*,u_*) be a local minimum of constant radius R for (P), where Φ is autonomous, ϕ continuously differentiable, and u_* bounded. Let f, Λ satisfy $\mathrm{L}_*^{\epsilon,R}$, with k_x^f, k_x^Λ, k_u^f, k_u^Λ summable. Then, if M_*^0 holds, there exist p and λ_0 satisfying the nontriviality and transversality conditions of Theorem 2.1, together with a summable function $\lambda(t)$ with values in $N_\Phi^C(\phi_t(x_*(t),u_*(t)))$* a.e. *such that the following adjoint inclusion holds:*

$$(-p'(t),0) \in \partial_C\big\{\, \langle p(t),f_t\rangle - \lambda_0\Lambda_t - \langle\lambda(t),\phi_t\rangle \,\big\}(x_*(t),u_*(t)) \text{ a.e.}$$

Furthermore , we have a local Weierstrass condition: for some $\delta > 0$, for t a.e.,

$$\begin{aligned}&|u-u_*(t)| \le \delta,\ \phi_t(x_*(t),u) \in \Phi \implies \\ &\langle p(t),f_t(x_*(t),u)\rangle - \lambda_0\Lambda_t(x_*(t),u) \ \le\ \langle p(t),f_t(x_*(t),u_*(t))\rangle - \lambda_0\Lambda_t(x_*(t),u_*(t)).\end{aligned}$$

Proof. Since ϕ is C^1 and u_* is bounded, ϕ satisfies $\mathrm{L}_*^{\epsilon,R}$ with constant k_x^ϕ and k_u^ϕ. Mimicking the proof of Prop. 4.6, we show that $\mathrm{M}_*^{\eta,\delta}$ holds for some $\eta \in (0,\epsilon)$ and $\delta \in (0,R)$ sufficiently small, with a constant M (we omit the details). We then apply Theorem 4.3 to directly obtain the stated conclusions. □

5. The multiplier rule in the calculus of variations. We consider in this section the historically significant question of formulating a multiplier rule in the calculus of variations, in the context of the following classical *problem of Lagrange*:

$$\text{minimize} \quad J(x) \;:=\; \ell(x(a),x(b)) + \int_a^b \Lambda_t(x(t),x'(t))\,dt$$

over the arcs x satisfying the following boundary conditions and pointwise constraint:

$$(x(a),x(b)) \in E,\; h_t(x(t),x'(t)) = 0 \text{ a.e. } t \in [a,b],$$

where $h : [a,b] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^N$. There is a large literature on the issue: see Hestenes [23] and the references therein. We demonstrate in this section how Theorem 2.1 subsumes and extends the known results. The above problem is the special case of (P) of Section 2 in which $f_t(x,u) = u$, so that u is identified with x'. Here, the set $S(t) = \{(x,u) : h_t(x,u) = 0\}$. We consider as before a local minimum x_* of radius R for the problem; that is, for any admissible arc x satisfying

$$|x(t) - x_*(t)| \leq \epsilon,\; |x'(t) - x_*'(t)| \leq R(t) \text{ a.e.}, \int_a^b |x'(t) - x_*'(t)|\,dt \leq \epsilon,$$

we have $J(x_*) \leq J(x)$. The Lipschitz hypothesis $\mathrm{L}_*^{\epsilon,R}$ for Λ is retained (it is evidently satisfied by f with $k_x^f = 0$, $k_u^f = 1$). We assume that h is $\mathcal{L}\times\mathcal{B}$ measurable and also satisfies $\mathrm{L}_*^{\epsilon,R}$. We posit moreover the following:

$\mathrm{M}_*^{\epsilon,R}$: There exists $M : [a,b] \to \mathbb{R}$ measurable such that, for almost every t,

$$(x,u) \in S_*^{\epsilon,R}(t),\; \lambda \in \mathbb{R}^N,\; (\alpha,\beta) \in \partial_L \langle \lambda, h_t\rangle(x,u) \implies |\lambda| \leq M(t)|\beta|.$$

It is easy to see that this is precisely the condition $\mathrm{M}_*^{\epsilon,R}$ of Section 4 if we identify ϕ with h and Φ with $\{0\}$ (which is the case under consideration).

THEOREM 5.1. *Under the hypotheses above, let the functions k_x^Λ and $Mk_x^h(1+k_u^\Lambda)$ be summable, and suppose that for some $\eta > 0$ we have $R(t) \geq \eta M(t) k_x^h(t)$ a.e. Then there exist p and λ_0 satisfying all the conclusions of Theorem 2.1, where the Euler adjoint inclusion is expressible in the form:*

$$(p'(t),p(t)) \in \partial_C\big\{\lambda_0 \Lambda_t\big\}(x_*(t),x_*'(t)) + N_{S(t)}^C(x_*(t),x_*'(t)) \text{ a.e.}$$

Furthermore, if h_t is strictly differentiable at $(x_*(t),x_*'(t))$ a.e., then there exists a measurable function $\lambda : [a,b] \to \mathbb{R}^N$ with $|\lambda(t)| \leq M(t)\{|p(t)| + \lambda_0 k_u^\Lambda(t)\}$ a.e. such that the Euler inclusion takes the explicit multiplier form

$$(p'(t),p(t)) \in \partial_C\big\{\lambda_0\Lambda_t + \langle \lambda(t), h_t\rangle\,\big\}(x_*(t),x_*'(t)) \text{ a.e.}$$

In the case of a sequence of radius functions R_i going to $+\infty$, the given bound on $|\lambda(t)|$ holds for the data M, k_u^Λ corresponding to any one of the radius functions.

Proof. With the identifications that have been made, this is a special case of Theorem 4.3; the Euler inclusion takes the form given here when $f_t(x,u) \equiv u$. ☐

We conclude this section with a brief discussion that clarifies how the theorem above subsumes the multiplier rules found in the classical literature, where x_* is Lipschitz and where L, h are continuously differentiable (at least). For any finite constant R, the regularity of L and the boundedness of x_*' provide the requisite

Lipschitz behavior $\mathrm{L}_*^{\epsilon,R}$ (with constant Lipschitz functions). A *rank hypothesis* is made, in which it is assumed that the matrix $D_u h_t(x,u)$ is of maximal rank N at points for which $h_t(x,u) = 0$. Since this implies MFC, Proposition 4.6 shows that this furnishes the constraint qualification required by Theorem 5.1, with a constant M (whence the required summability). In the case of a strong local minimum (a local $W^{1,1}$ minimum actually suffices), the theorem allows us to let $R \to +\infty$ in order to obtain the global Weierstrass condition. It follows in this classical setting that the resulting multiplier $\lambda(t)$ is *essentially bounded* (since, for the first radius (say), M and k_u^Λ are constants). We summarize the result for the classical setting:

COROLLARY 5.2. *Let the Lipschitz arc x_* provide a strong local minimum for $J(x)$ subject to the boundary conditions and the constraint $h_t(x,x') = 0$ a.e., where $L, h \in C^1$, and where, for some $\epsilon > 0$, $D_u h$ is of rank N at all points (t,x,u) for which $|x - x_*(t)| < \epsilon$ and $h_t(x,u) = 0$. Then the conclusions of Theorem 5.1 hold, with a global Weierstrass condition, and for a bounded multiplier $\lambda(t)$.*

We remark that in some classical variants of the multiplier rule, it is assumed that x_*' is piecewise continuous and that the rank hypothesis holds solely along x_*, in the following sense: $D_u h(t, x_*(t), x_*'(t))$ has maximal rank for each t, where at a corner, this is taken to hold for both $x_*'(t+)$ and $x_*'(t-)$. That situation is covered by Theorem 4.8.

6. Differential-algebraic systems. In this section we consider the following problem, for a given function $h : [a,b] \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^N$: to minimize

$$J(x,u) := \ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), u(t), x'(t))\, dt$$

over the pairs (x,u) satisfying the controlled differential-algebraic system (or implicit differential equation) and control constraints

$$h_t(x(t), u(t), x'(t)) = 0,\ u(t) \in U(t) \text{ a.e.}$$

as well as the boundary conditions $(x(a), x(b)) \in E$. The problem is viewed as a special case of (P) of Section 2 in which the control variable is (u,v), $f_t(x,u,v) := v$, and

$$S := \big\{(t,x,u,v) : t \in [a,b], u \in U(t),\ h_t(x,u,v) = 0\big\}.$$

For a given admissible process (x_*, u_*, v_*), radius function R, and $\epsilon > 0$, we suppose as before that (x_*, u_*, v_*) is a local minimum of radius R for the problem. We take U and h $\mathcal{L}\times\mathcal{B}$ measurable and U closed-valued, and we suppose that h and Λ satisfy $\mathrm{L}_*^{\epsilon,R}$; evidently, f satisfies $\mathrm{L}_*^{\epsilon,R}$ with $k_x^f = 0 = k_u^f$ and $k_v^f = 1$. The following *constraint qualification* is made:

$\mathbf{M}_*^{\epsilon,R}$ **:** There exists $M : [a,b] \to \mathbb{R}$ measurable such that, for almost every t,

$$(x,u,v) \in S_*^{\epsilon,R}(t),\ \lambda \in \mathbb{R}^N,\ \mu \in N_{U(t)}^L(u),\ (\alpha, \beta - \mu, \gamma) \in \partial_L \langle \lambda, h_t \rangle (x,u,v) \implies |\lambda| \le M(t)|(\beta,\gamma)|.$$

It is easy to see that this coincides with the condition $\mathrm{M}_*^{\epsilon,R}$ of Section 4, when the choice $\phi_t = [\, h_t(x,u,v), u\,]$, $\Phi(t) = \{0\} \times U(t)$ is made (which is the case here). With these identifications, the following is immediate from Theorem 4.3:

THEOREM 6.1. *Under the hypotheses above, let the functions* k_x^Λ, $Mk_x^h[1+k_u^\Lambda+k_v^\Lambda]$ *be summable, and suppose that for some* $\eta > 0$ *we have* $R(t) \geq \eta M(t) k_x^h(t)$ a.e. *Then there exist* p *and* λ_0 *satisfying all the conclusions of Theorem 4.3 where, in the explicit multiplier case, the adjoint inclusion is expressible in the form*

$$(p'(t), -\mu(t), p(t)) \in \partial_C\big\{\lambda_0 \Lambda_t + \langle \lambda(t), h_t\rangle\,\big\}(x_*(t), u_*(t), x_*'(t)) \text{ a.e.}$$

for a measurable function $\mu(\cdot)$ *taking values in* $N^C_{U(t)}(u_*(t))$ a.e.*, and where*

$$|\lambda(t)| \leq M(t)\big\{|p(t)| + \lambda_0 k_u^\Lambda(t) + \lambda_0 k_v^\Lambda(t)\big\} \text{ a.e.}$$

We conclude this section with a special case that facilitates comparison to the existing literature, which has relied upon implicit function theorems and more restrictive hypotheses. Let Λ, h be continuously differentiable and U autonomous, and assume that u_* is bounded. In this context, it follows that for any constant finite radius R, the functions k_x^h, k_x^Λ, k_u^Λ, k_v^Λ may all be taken to be constant. With ϕ and Φ defined as above, the condition MFC of Section 4 at a point (t, x, u, v) reduces to the following:

$$(0,0) \in D_{u,v}\langle \lambda, h_t(x,u,v)\rangle + N_U^L(u)\times\{0\} \Longrightarrow \lambda = 0, \tag{6.1}$$

and Prop. 4.6 implies that if (6.1) is satisfied on $C_*^{\epsilon,R}$, then $\mathrm{M}_*^{\epsilon,R}$ holds (with a constant M). By considering a sequence $R_i \to +\infty$, we obtain the following:

COROLLARY 6.2. *For* $\epsilon > 0$*, let* (x_*, u_*) *provide a minimum for the problem relative to the admissible* (x,u) *satisfying* $|x(t) - x_*(t)| < \epsilon$*, where* Λ, h *are* C^1*,* U *is autonomous, and* u_* *is bounded. Suppose that* (6.1) *holds at each point* $(t,x,u,v) \in S$ *satisfying* $|x - x_*(t)| < \epsilon$*. Then the conclusions of the Theorem hold, with the global Weierstrass condition, and for a multiplier* $\lambda(t)$ *that is bounded.*

7. Constraints of equality/inequality type. In this section we consider the following problem: to minimize

$$J(x,u) := \ell(x(a), x(b)) + \int_a^b \Lambda_t(x(t), u(t))\, dt$$

subject to the dynamics, unilateral control constraints, and boundary conditions

$$x'(t) = f_t(x(t), u(t)),\ u \in U(t) \text{ a.e.},\ (x(a), x(b)) \in E,$$

as well as the pointwise mixed constraints

$$g_t(x(t), u(t)) \leq 0,\ h_t(x(t), u(t)) = 0 \text{ a.e.}$$

Here h has values in $\mathbb{R}^N$ and g in $\mathbb{R}^K$ (so that the inequality $g \leq 0$ is understood in the vector sense). This type of context has dominated the literature on mixed constraints. The problem may be viewed as the special case of (P) of Section 2 in which

$$S := \big\{(t,x,u) : t \in [a,b], u \in U(t),\ g_t(x,u) \leq 0,\ h_t(x,u) = 0\big\}.$$

For a given admissible process (x_*, u_*), radius function R, and $\epsilon > 0$, we suppose as before that (x_*, u_*) is a local minimum of radius R for the problem. We take U, f,

g, h, Λ measurable, U closed-valued, and we suppose that the functions f, g, h, Λ satisfy $\mathrm{L}_*^{\epsilon,R}$.

In the absence of the set U (or at points in its interior), the classical (smooth and uncalibrated) Mangasarian-Fromowitz condition MFC corresponding to our equality/inequality system is given by

$$\lambda \in \mathbb{R}^N,\ \gamma \in \mathbb{R}_+^K,\ \langle \gamma, g_t(x,u)\rangle = 0,\ D_u\big\{ \langle \gamma, g_t\rangle + \langle \lambda, h_t\rangle \big\}(x,u) = 0 \Longrightarrow \gamma = 0,\ \lambda = 0.$$

This is equivalent to requiring that, at admissible points, the gradients of the functions $\{D_u g_i, D_u h_j\}$ (where i is restricted to those indices for which the inequality constraint is saturated) are linearly independent, with the further proviso that only nonnegative coefficients need be considered for the $D_u g_i$. When U is present, the condition reads

$$\lambda \in \mathbb{R}^N,\ \gamma \in \mathbb{R}_+^K,\ \langle \gamma, g_t(x,u)\rangle = 0,$$
$$D_u\big\{ \langle \gamma, g_t\rangle + \langle \lambda, h_t\rangle \big\}(x,u) \in -N^L_{U(t)}(u) \Longrightarrow \gamma = 0,\ \lambda = 0.$$

The following calibrated and nonsmooth extension of this constraint qualification is imposed:

$\mathbf{M}_*^{\epsilon,R}$ **:** There exists $M : [a,b] \to \mathbb{R}$ measurable such that, for almost every t,

$$(x,u) \in S_*^{\epsilon,R}(t),\ \lambda \in \mathbb{R}^N,\ \mu \in N^L_{U(t)}(u),\ \gamma \in \mathbb{R}_+^K,\ \langle \gamma, g_t(x,u)\rangle = 0,$$
$$(\alpha, \beta - \mu) \in \partial_L\big\{ \langle \gamma, g_t\rangle + \langle \lambda, h_t\rangle \big\}(x,u) \Longrightarrow |(\gamma,\lambda)| \le M(t)|\beta|.$$

This coincides with the condition $\mathrm{M}_*^{\epsilon,R}$ of Section 4 for the relevant identifications:

$$\phi_t(x,u) = [\, g_t(x,u), h_t(x,u), u\,],\ \Phi(t) = \mathbb{R}_-^K \times \{0\} \times U(t).$$

Thus the validity of $\mathrm{M}_*^{\epsilon,R}$ (for a constant M) would follow in certain scenarios from Propositions 4.5, 4.4, or 4.6. As a special case of Theorem 4.3 we obtain the following *intermediate* theorem (that is, for a given radius function R):

THEOREM 7.1. *Under the hypotheses above, let the functions*

$$k_x^f,\ k_x^\Lambda,\ M\big\{k_x^g + k_x^h\big\}\big\{k_u^f + k_u^\Lambda\big\}$$

be summable, and suppose that for some $\eta > 0$ we have

$$R(t)\ \ge\ \eta M(t)\big\{k_x^g(t) + k_x^h(t)\big\} \text{ a.e.}$$

Then there exist p and λ_0 satisfying all the conclusions of Theorem 2.1. If g_t, h_t are strictly differentiable at $(x_*(t), u_*(t))$ a.e.*, then there exist measurable functions*

$$\lambda : [a,b] \to \mathbb{R}^N,\ \gamma : [a,b] \to \mathbb{R}_+^K,\ \textit{with } \langle \gamma(t), g_t(x_*(t), u_*(t))\rangle = 0 \text{ a.e.}$$

and

$$|(\gamma(t), \lambda(t))|\ \le\ M(t)\big\{|p(t)|k_u^f(t) + \lambda_0 k_u^\Lambda(t)\big\} \text{ a.e.}$$

such that the adjoint inclusion is expressible in the explicit multiplier form

$$(-p'(t), \mu(t)) \in \partial_C\big\{ \langle p(t), f_t\rangle - \lambda_0\Lambda_t - \langle \gamma(t), g_t\rangle - \langle \lambda(t), h_t\rangle \big\}(x_*(t), u_*(t)) \text{ a.e.,}$$

where μ is a measurable function satisfying $\mu(t) \in N^C_{U(t)}(u_*(t))$ a.e.

We remark that this result, like the ones below, has an evident variant in which either the equality or inequality constraints are absent; the elements corresponding to h or g are simply deleted from the hypotheses and conclusions.

The intermediate form above generates, for example, the following *global* case of the theorem in which the data are smooth and the Mangasarian-Fromowitz condition appears in its classical form; the theorem results from invoking Prop. 4.6, and letting $R_i \to \infty$:

Corollary 7.2. *Let (x_*, u_*) be a strong (or $W^{1,1}$) local minimum, where all the functions involved are C^1, U is autonomous and u_* bounded. Suppose that for every t, at each point $(x, u) \in S(t)$ for which $|x - x_*(t)| < \epsilon$, the following constraint qualification holds:*

$$\lambda \in \mathbb{R}^N,\ \gamma \in \mathbb{R}^K_+,\ \langle \gamma, g_t(x,u) \rangle = 0,$$
$$D_u \big\{ \langle \gamma, g_t \rangle + \langle \lambda, h_t \rangle \big\}(x,u) \in -N^L_U(u) \implies \gamma = 0,\ \lambda = 0.$$

Then the conclusions of Theorem 7.1 hold, with the global Weierstrass condition.

Having considered the intermediate and global situations above, we now present a *local* (or pointwise) version, in which the constraint qualification is imposed only with reference to the optimal process itself. Theorem 4.8 yields:

Corollary 7.3. *We posit the same hypotheses as in Corollary 7.2, but with the constraint qualification assumed to hold only at admissible cluster points (t, x, u) of (x_*, u_*) (see Def. 4.7). Then there exist p, λ_0, λ, γ, μ satisfying the conclusions of Theorem 7.1, but with a local Weierstrass condition: for some $\delta > 0$, for t* a.e.,

$$(x_*(t), u) \in S(t),\ |u - u_*(t)| \le \delta \implies$$
$$\langle p(t), f_t(x_*(t), u) \rangle - \lambda_0 \Lambda_t(x_*(t), u) \ \le\ \langle p(t), f_t(x_*(t), u_*(t)) \rangle - \lambda_0 \Lambda_t(x_*(t), u_*(t)).$$

Another variation on our theme involves the use of Theorem 3.2 to allow asymmetric hypotheses relative to a partitioned control $u = (v, w)$. We illustrate this via the problem of minimizing the same cost functional $J(x,u) = J(x,v,w)$ as above, subject to the same dynamics $x' = f_t(x,v,w)$ and boundary conditions, but under the constraints

$$g_t(x(t), v(t), w(t)) \le 0,\ h_t(x(t), v(t), w(t)) = 0,\ w(t) \in W(t) \text{ a.e.},$$

where W is $\mathcal{L} \times \mathcal{B}$ measurable. We are given $u_* = (v_*, w_*)$, a local minimum of radius R for the problem. The basic measurability hypotheses remain in force on f, g, h, Λ, which are now assumed to satisfy the asymmetric hypothesis $\mathrm{HBS}^{\epsilon,R}_*$ of Theorem 3.2 (thus, continuity in w is not assumed). We define

$$S(t,w) := \big\{ (x,v) : g_t(x,v,w) \le 0,\ h_t(x,v,w) = 0 \big\}.$$

The following calibrated constraint qualification is imposed:

$\mathbf{HM}^{\epsilon,R}_*$ **:** There exists $M : [a,b] \to \mathbb{R}$ measurable such that, for almost every t,

$$w \in W(t) \cap B(w_*(t), R(t)),\ (x,v) \in S^{\epsilon,R}_*(t,w),\ \lambda \in \mathbb{R}^N,\ \gamma \in \mathbb{R}^K_+,\ \langle \gamma, g_t(x,v,w) \rangle = 0,$$
$$(\alpha, \beta) \in \partial_L \big\{ \langle \gamma, g_t(\cdot,\cdot,w) \rangle + \langle \lambda, h_t(\cdot,\cdot,w) \rangle \big\}(x,v) \implies |(\gamma,\lambda)| \le M(t)|\beta|.$$

We invoke Theorem 3.2 (together with Propositions 4.1 and 4.2) to derive:

THEOREM 7.4. *Under the hypotheses above, let the functions*

$$k_x^f(t, w_*(t)),\ \ k_x^\Lambda(t, w_*(t)) \text{ and}$$
$$M(t)\big\{k_v^f(t, w_*(t)) + k_v^\Lambda(t, w_*(t))\big\}\big\{k_x^g(t, w_*(t)) + k_x^h(t, w_*(t))\big\}$$

be summable, and suppose that for some $\eta > 0$ we have

$$R(t) \ \geq\ \eta M(t)\big\{k_x^g(t, w_*(t)) + k_x^h(t, w_*(t))\big\} \text{ a.e.}$$

Then there exist p and λ_0 satisfying all the conclusions of Theorem 3.2. If $g_t(\cdot, \cdot, w_*(t))$ and $h_t(\cdot, \cdot, w_*(t))$ are strictly differentiable at $(x_*(t), v_*(t))$ a.e., then there exist measurable functions

$$\lambda : [a, b] \to \mathbb{R}^N,\ \gamma : [a, b] \to \mathbb{R}_+^K,\ \textit{with}\ \langle \gamma(t), g_t(x_*(t), u_*(t))\rangle = 0 \text{ a.e.}$$

and

$$|(\gamma(t), \lambda(t))| \ \leq\ M(t)\big\{|p(t)|k_v^f(t, w_*(t)) + \lambda_0 k_v^\Lambda(t, w_*(t))\big\} \text{ a.e.}$$

such that the hybrid adjoint inclusion is expressible in the explicit multiplier form

$$(-p'(t), 0) \in \partial_C\big\{\ \langle p(t), f_t\rangle - \lambda_0\Lambda_t - \langle \gamma(t), g_t\rangle - \langle \lambda(t), h_t\rangle\ \big\}(x_*(t), v_*(t), w_*(t)) \text{ a.e.},$$

the generalized gradient being taken with respect to (x, v).

8. Relation to the literature. This section relates our results to the existing literature. It is not meant to be a comprehensive survey, nor do we go beyond the current framework: first-order necessary conditions for optimal control problems with standard dynamics that include mixed constraints but not unilateral state constraints. (We believe, however, that the methods will extend to that case.)

One of the novel features of this article is the presence of the radius function $R(t)$. Let us briefly explain how this stratified aspect is new. Even in the classical calculus of variations, the addition of a constraint such as $|x'(t) - x_*'(t)| \leq R$ modifies the concept of local minimum. In a control setting, the question is whether the radius constraint $|u - u_*(t)| \leq R(t)$ can just be absorbed into the control constraint structure, thereby changing our local minimum into one of the usual kind for the redefined problem, and making it amenable to known results.

In the case of a standard optimal control problem with unmixed constraints $u \in U$, this reduction is possible: we simple replace U by $U \cap B(u_*(t), R(t))$, and known necessary conditions then apply. The case of mixed constraints is different, because of the fact that certain constraint qualifications (rank hypotheses, Mangasarian-Fromowitz or bounded slope conditions) must be satisfied before the necessary conditions can be invoked. To illustrate, consider the case in which the only control constraint is $h(x, u) = 0$. For a strong local minimum, in order to assert the necessary conditions, the usual multiplier rule requires the hypothesis that the rank of h_u be maximal when $h = 0$. Now suppose that we have instead a local minimum of radius R. In order to absorb this additional constraint into the set of admissible control processes, we would need to consider *two* constraints: $h(x, u) = 0$, $|u - u_*(t)| \leq R(t)$, where both are active at certain points. Now the required rank condition that would allow us to invoke the existing necessary conditions has changed: it is more complicated than h_u being of maximal rank. In our theorem, however, the radius constraint is part of the *definition* of local minimum, and our results show that it *does* suffice to have the h_u rank condition. This is new (and useful).

The role of the radius in obtaining directly either global, local, or pointwise versions of the necessary conditions is amply illustrated in Section 3 (in connection with the Euler adjoint inclusion), in Section 5 (in connection with the classical multiplier rule), and in Section 7 (in connection with equality/inequality constraints). In some cases, this leads to the affirmation of a Weierstrass condition in contexts where prior results made no such claim. The use of a radius also plays an important role in our forthcoming extension of the Schwarzkopf multiplier rule (see below).

Another innovative aspect of our results is the weak regularity hypotheses imposed on the data of the problem, for which continuous differentiability has been the norm (with a few exceptions, see below). The general nature of the mixed constraint is also new, whether it is the geometric form used in Theorem 2.1 or the functional form $\phi_t(x,u) \in \Phi_t$ studied in Section 4; prior results in the literature have been framed in terms of equalities and inequalities. Our results also inherit from earlier work on the nonsmooth maximum principle the very general nature of the endpoint constraints, which need not be prescribed by a smooth manifold, in contrast to most of the literature.

We now make more specific reference to related work. The best source for what could be termed the classical approach is undoubtedly Hestenes [23], which synthesizes the classical multiplier rules in the calculus of variations and extends them to control problems. We find there piecewise continuous controls and data that is continuously differentiable in all variables. There is no unilateral control set constraint to accompany the mixed constraints which, as is the case for all the references discussed below, are taken to be of equality/inequality type. A global rank hypothesis is made (as in Section 5), but the reduction of independence to *positive* linear independence for the inequality constraints (which is what the Mangasarian-Fromowitz condition would give in this context) is not made. Section 5 shows how these results are subsumed and extended.

Significant progress was made in the work of Neustadt and Makowski [26, 28]. The data remain continuously differentiable, but controls are bounded and measurable, and a unilateral control set that is 'open in itself' is allowed. The mixed constraints are not necessarily 'regular': no linear independence is imposed. In this context, the adjoint variable is a function of bounded variation.

In [20], Dmitruk presents the results of the Dubovitskiĭ-Milyutin school for the problem in question. This work features a partitioned control (v,w) as in Section 3, and a constraint qualification including positive linear independence of the inequality constraints. The constraint qualification, however, must be satisfied by the part of the control v that is *not* subject to a unilateral constraint, in contrast to our Theorem 7.1 or Corollary 7.2. The regularity hypotheses on the data are also significantly stronger than ours (see Theorem 7.4): bounded control, continuous differentiability in (x,v), and continuity in w; furthermore, the boundary conditions correspond to smooth sets. Unilateral state constraints are allowed, however.

In [19], Devdariani and Ledyaev break new ground by studying a difficult problem with implicit dynamics (as in Section 6), and nonsmooth mixed constraints. The analysis is limited to fixed initial condition and free endpoints, however, which is quite restrictive.

The article by de Pinho and Rosenblueth [15] also allows merely Lipschitz dynamics and cost, much like ours, and has the same general boundary conditions. Their constraint functions are smooth in the state and the control, however, and their results do not include a Weierstrass condition. A full rank hypothesis is posited in a partitioned framework, along the lines of [20], but only pointwise along the optimal process. This resembles our hypothesis in Theorem 4.8, but without the use of cluster points. The following example shows that cluster points cannot be ignored, however.

An example. In one dimension (for x) and two dimensions (for u), on the interval $[0,1]$, we minimize $-x(1)$ subject to $x'(t) = u_1(t)$, $x(0) = -1$ and the mixed constraint $x(t)+u_2(t) \le 0$ a.e., where u_1, u_2 are constrained by $|u_i| \le 2$ and $u_1u_2 \ge 0$. It is clear that any admissible trajectory has $x(t) \le 0 \;\forall\, t$, for otherwise there would exist a set of positive measure in which $x(t)$ and $x'(t)$ are strictly positive, whence $u_1(t)u_2(t) < 0$, a contradiction. Thus the cost corresponding to any admissible trajectory is always nonnegative, and the arc $x_*(t) = t - 1$ is optimal, for the control

$$(u_1(t), u_2(t)) = \begin{cases} (1,0) & \text{if } t \in [0,1) \\ (0,-1) & \text{if } t = 1 \end{cases}$$

Note that the single inequality constraint is unsaturated at every t, and there is no equality constraint. Thus MFC holds at every point along the optimal process. This may lead us to believe that the necessary conditions that apply in the absence of the inequality constraint (the usual maximum principle) hold; indeed, this is asserted in [15]. Such is not the case, however. (The necessary conditions would give $p \equiv 0$ and (by transversality) $\lambda_0 = 0$, contradicting nontriviality.) The inapplicability of our Corollary 7.3 is due to the fact that the constraint qualification fails (as it must), in this case at a single cluster point $(t, x, u_1, u_2) = (1, 0, 1, 0)$: we have there

$$D_u\big\{\gamma(x+u_2)\big\}(1,0,1,0) = (0,\gamma) \in -N_U^L(1,0) = \{0\} \times \mathbb{R}_+$$

for any $\gamma > 0$. We remark that when only equality constraints are present, the results of [15] are correct and deducible from ours: one may show that the hypotheses there (which are of the type that appear in Prop. 4.4) imply that $\mathrm{M}_*^{\epsilon,R}$ holds for a constant M, when R is a sufficiently small constant. Thus we may add the Weierstrass condition to the list of conclusions.

Finally, we note the existence of a multiplier rule due to Schwarzkopf [33] which posits no explicit constraint qualification, but instead imposes a local surjectivity condition on the functions defining the mixed constraint. (A partial convexity hypothesis is also required.) This result, which is also unusual in asserting a Weierstrass condition involving nonadmissible points, has long been viewed as an anomaly in the theory. It turns out, however, that it can be obtained (and extended) by means of Theorem 2.1, though not as directly as the others we have discussed. The details are given in a companion paper [9].

9. Proofs of Theorems 2.1 and 3.2. We shall first prove the theorem in the case $\Lambda \equiv 0$. We may assume without loss of generality that k_x^f, k_S are positive-valued. Set

$$k(t) := k_x^f(t) + k_S(t)k_u^f(t) \in L^1(a,b),\; c(t) := Mk(t)/k_S(t),$$

where $M > 1$ is a parameter, and define

$$F(t,x) := \{(f_t(x,u), \theta_t(u)) : u \in S(t,x)\}\,,\; t \in [a,b], |x - x_*(t)| \le \epsilon,$$

where $\theta_t(u) := c(t)(u - u_*(t))$. The graph $G(t)$ of $F(t,\cdot)$ is then defined as usual:

$$G(t) := \{(x, f(t,x,u), \theta_t(u)) : (x,u) \in S(t)\}\,.$$

It follows that F is $\mathcal{L}\times\mathcal{B}$ measurable, and that for almost every t, $G(t)$ is locally closed around points $(x, f_t(x,u), \theta_t(u))$ for which $(x,u) \in S_*^{\epsilon,R}(t)$, $|u-u_*(t)| < R(t)$. We set

$$K(t) := 2k_x^f(t) + (1+k_S(t))k_u^f(t),\ K_M'(t) := Mk(t)(1+k_S(t))/k_S(t).$$

PROPOSITION 9.1. *Let (α,β,τ) belong to $N_{G(t)}^L(x, f_t(x,u), \theta_t(u))$, where $(x,u) \in S_*^{\epsilon,R}(t)$ and $|u-u_*(t)| < R(t)$. Then for almost every t, we have*

$$|\alpha| \ \le\ k(t)\left\{|\beta| + M|\tau|\right\}, \tag{9.1}$$

as well as $(\alpha, 0) \in \partial_L\big\{\langle -\beta, f_t\rangle - \langle \tau, \theta_t\rangle + [K(t)|\beta| + K_M'(t)|\tau|]d_{S(t)}\big\}(x,u)$.

Proof. We suppose that t is a point where the properties in $\mathrm{L}_*^{\epsilon,R}$ and $\mathrm{BS}_*^{\epsilon,R}$ hold. We suppress t in order to simplify the notation, and we first treat the case in which the following stronger inclusion holds: $(\alpha,\beta,\tau) \in N_G^P(x, f(x,u), \theta(u))$. By the definition of proximal normal, there exists $\sigma \ge 0$ such that the function $(x',u') \mapsto$

$$\langle -(\alpha,\beta,\tau), (x', f(x',u'), \theta(u'))\rangle + \sigma\left\{|(x', f(x',u'), \theta(u')) - (x, f(x,u), \theta(u))|\right\}^2$$

has a local minimum relative to $(x',u') \in S$ at $(x',u') = (x,u)$. From the Lipschitz properties of f and θ it follows that for a sufficiently large value of σ', the function

$$h(x',u') := \langle -(\alpha,\beta,\tau), (x', f(x',u'), \theta(u'))\rangle + \sigma'\left\{|x'-x|^2 + |u'-u|^2\right\}$$

also has a local minimum relative to $(x',u') \in S$ at $(x',u') = (x,u)$. Writing the necessary condition $(0,0) \in \partial_L\{h + I_S\}(x,u)$ for the minimum (where I_S is the indicator function of S) yields

$$(\alpha, 0) \in \partial_L\left\{\langle -\beta, f(x,u)\rangle - \langle \tau, \theta(u)\rangle\right\} + N_S^L(x,u),$$

by the sum rule [10, 1.10.1], and since $\partial_L I_S = N_S^L$. In light of this inclusion, there exists

$$(\alpha', \beta') \in \partial_L\left\{\langle -\beta, f(x,u)\rangle - \langle \tau, \theta(u)\rangle\right\}$$

such that $(\alpha - \alpha', -\beta') \in N_S^L(x,u)$. We deduce from $\mathrm{BS}_*^{\epsilon,R}$ the inequality $|\alpha - \alpha'| \le k_S|\beta'|$, since $N_S^L(x,u)$ is generated by limits from $N_S^P(x,u)$. Furthermore, (α', β') satisfies $|\alpha'| \le k_x^f|\beta|$, $|\beta'| \le k_u^f|\beta| + c|\tau|$, in view of the Lipschitz condition. These facts yield

$$|\alpha| \le |\alpha'| + k_S|\beta'| \le k_x^f|\beta| + k_S(k_u^f|\beta| + c|\tau|) = (k_x^f + k_S k_u^f)|\beta| + k_S c|\tau|,$$

which gives rise to the estimate (9.1) claimed in the Proposition. It follows from this estimate that for any $m > 0$, the function h admits the Lipschitz constant $K(t)|\beta| + K_M'(t)|\tau| + m$ in a sufficiently small neighborhood of (x,u), where K, K_M' were defined above. An exact penalization result [5, Prop. 2.4.3] implies then that the function $h + [K(t)|\beta| + K_M'(t)|\tau| + m]d_S$ attains a local minimum at (x,u), whence

$$(0,0) \in \partial_P\left\{h + [K(t)|\beta| + K_M'(t)|\tau| + m]d_S\right\}(x,u), \text{ or}$$

$$\begin{aligned}(\alpha, 0) &\in \partial_L\left\{\langle -\beta, f(x,u)\rangle - \langle \tau, \theta(u)\rangle + [K(t)|\beta| + K_M'(t)|\tau| + m]d_S(x,u)\right\}\\ &\subset \partial_L\left\{\langle -\beta, f(x,u)\rangle - \langle \tau, \theta(u)\rangle + [K(t)|\beta| + K_M'(t)|\tau|]d_S(x,u)\right\} + m\partial_L d_S(x,u).\end{aligned}$$

Letting $m \downarrow 0$ gives the desired result.

Consider now the general case, in which $(\alpha, \beta, \tau) \in N_G^L(x, f(x,u), \theta(u))$. Then there exist sequences $(\alpha_i, \beta_i, \tau_i)$ converging to (α, β, τ) and $(x_i, f(x_i, u_i), \theta(u_i))$ converging to $(x, f(x,u), \theta(u))$ such that

$$(\alpha_i, \beta_i, \tau_i) \in N_G^P(x_i, f(x_i, u_i), \theta(u_i)).$$

The estimate (9.1) for $|\alpha_i|$ proven above (in terms of β_i and τ_i) clearly implies the required one in the limit. There remains the inclusion to verify.

It follows from the nature of θ that $u_i \to u$, and (by the case treated above) that

$$\begin{aligned}
(\alpha_i, 0) &\in \partial_L \left\{ \langle -\beta_i, f(x_i, u_i) \rangle - \langle \tau_i, \theta(u_i) \rangle + [K(t)|\beta_i| + K_M'(t)|\tau_i|] d_S(x_i, u_i) \right\} \\
&\subset \partial_L \left\{ \langle -\beta, f(x_i, u_i) \rangle - \langle \tau, \theta(u_i) \rangle + [K(t)|\beta| + K_M'(t)|\tau|] d_S(x_i, u_i) \right\} \\
&\quad + \partial_L \left\{ \langle \beta - \beta_i, f(x_i, u_i) \rangle - \langle \tau_i - \tau, \theta(u_i) \rangle \right\} \\
&\quad + \partial_L \left\{ [K(t)(|\beta_i| - |\beta|) + K_M'(t)(|\tau_i| - |\tau_i|)] d_S(x_i, u_i) \right\}.
\end{aligned}$$

Passing to the limit in this relation, we obtain the required conclusion. □

An auxiliary problem. We now define $y_*(t) \equiv 0$ as well as a radius function $R_F(t) := c(t)R(t)$ (this is naturally taken to be $+\infty$ when $R(t) = +\infty$). We proceed to observe that the arc (x_*, y_*) is admissible for the differential inclusion optimal control problem (Q) of minimizing $\ell(x(a), x(b))$ over the arcs (x, y) satisfying

$$|(x'(t), y'(t)) - (x_*'(t), y_*'(t))| \le R_F(t), \ |x(t) - x_*(t)| \le \epsilon \text{ a.e.}, \ \int_a^b |x'(t) - x_*'(t)| \, dt \le \epsilon$$

as well as

$$(x(a), x(b)) \in E, \quad y(a) = 0, \quad (x'(t), y'(t)) \ \in \ F(t, x(t)) \text{ a.e.}$$

Indeed, (x_*, y_*) is a solution of this problem. This results from the fact that any trajectory (x, y) for F as above admits a measurable $u(\cdot)$ such that $(x(t), u(t)) \in S(t)$ and $(x'(t), y'(t)) = (f_t(x(t), u(t)), \theta_t(u(t)))$ a.e. (this uses a measurable selection theorem). Furthermore, we have, for almost every t,

$$|y'(t) - y_*'(t)| = |\theta_t(u(t))| = c(t)|u(t) - u_*(t)| \le R_F(t),$$

so that $|u(t) - u_*(t)| \le R_F(t)/c(t) = R(t)$. But we know that x_* provides a minimum of radius R for the problem (P), and it follows that (x_*, y_*) is optimal for (Q).

We now bring to bear upon (Q) Corollary 3.5.3 of [7], whose hypotheses we verify. We have seen that (x_*, y_*) is a local minimum of radius R_F for this problem. The condition (9.1) of Proposition 9.1 confirms the bounded slope condition of radius R_F required by that result, with the function Mk. We also have the ratio $R_F(t)/(Mk(t)) = R(t)/k_S(t) \ge \eta$ bounded away from 0 as required.

We deduce the existence of an arc (p, q) and a number λ_0 in $\{0, 1\}$ satisfying the nontriviality condition

$$(\lambda_0, p(t), q(t)) \ne 0 \ \ \forall \, t \in [a, b]$$

and the transversality condition:

$$(p(a), -p(b)) \in \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N_E^L(x_*(a), x_*(b)), \ q(b) = 0$$

and such that p satisfies the Euler adjoint inclusion:

$$(-p'(t), -q'(t)) \in \\ \text{co}\left\{(\omega,\nu) : (\omega,\nu,p(t),q(t)) \in N^L_{G^+(t)}(x_*(t), y_*(t), f_t(x_*(t),u_*(t)),0)\right\},$$

where $G^+(t)$ is the set $\{(x,y,f_t(x,u),\theta_t(u)) : (x,u) \in S(t)\}$. Because this set imposes no restrictions on y, it follows that any point $(\omega,\nu,p(t),q(t))$ as above has $\nu = 0$, whence $q'(t) = 0$ a.e., so that $q(t) \equiv 0$ (since $q(b) = 0$).

It is a further conclusion that (p,q) satisfies the Weierstrass condition of radius R_F for almost every t : if $u \in S(t, x_*(t))$ satisfies the condition

$$|(f_t(x_*(t),u), \theta_t(u)) - (f_t(x_*(t),u_*(t)),0)| \le R_F(t), \text{ then} \tag{9.2}$$

$$\langle p(t), f_t(x_*(t),u)\rangle + \langle q(t), \theta_t(u)\rangle \le \langle p(t), f_t(x_*(t),u_*(t))\rangle + \langle q(t), \theta_t(u_*(t))\rangle\,,$$

or, since $q = 0$, $\langle p(t), f_t(x_*(t),u)\rangle \le \langle p(t), f_t(x_*(t),u_*(t))\rangle$. But (in view of the Lipschitz condition) (9.2) is certainly satisfied if

$$|u - u_*(t)| \le \frac{R_F(t)}{k^f_u(t) + c(t)} = \frac{c(t)R(t)}{k^f_u(t) + c(t)} = \frac{M(k(t)/k_S(t))R(t)}{k^f_u(t) + Mk(t)/k_S(t)}.$$

This last expression is no less than

$$\frac{Mk(t)R(t)}{k^f_x + k_S(t)k^f_u(t) + Mk(t)} = \frac{M}{M+1}R(t).$$

We obtain therefore the following Weierstrass condition of radius $RM/(M+1)$:

$$t \text{ a.e.}, u \in S(t,x_*(t)), |u - u_*(t)| \le R(t)M/(M+1) \\ \Longrightarrow \langle p(t), f_t(x_*(t),u)\rangle \le \langle p(t), f_t(x_*(t),u_*(t))\rangle\,.$$

Bearing in mind again that $q \equiv 0$, we obtain $(\lambda_0, p(t)) \ne 0$ on $[a,b]$, as well as

$$(p'(t),0) \in \text{co}\left\{(\omega,0) : (\omega,p(t),0) \in N^L_{G(t)}(x_*(t), f(t,x_*(t),u_*(t)),0)\right\}. \tag{9.3}$$

We now produce an arc p that satisfies all the above, including the Weierstrass condition of full radius R, by allowing $M \to \infty$. This is carried out by a standard argument that exploits the fact that the set of (λ_0, p) satisfying transversality, the adjoint inclusion (9.3) and $\lambda_0 + \|p\|_\infty = 1$ (with $\lambda_0 \ge 0$) is sequentially compact (see [7, Prop. 2.1.2]). We first get an arc p_i as above for each integer $M = i > 1$. By normalizing, we arrange to have $\lambda_{0,i} + \|p_i\|_\infty = 1$. Each p_i (as a consequence of (9.3) and (9.1)) satisfies $|p_i'(t)| \le k(t)|p_i(t)|$ a.e., a uniform estimate which (with the help of Gronwall's Lemma) allows us to extract a convergent subsequence of the $(\lambda_{0,i}, p_i)$ whose limit (λ_0, p) satisfies all the above, including the adjoint inclusion (9.3), but now with the Weierstrass condition of full radius R. We then renormalize if necessary to have $\lambda_0 \in \{0,1\}$.

At this point, we have an arc p satisfying all the required conclusions of Theorem 2.1, except that the adjoint inclusion is of the form (9.3) instead of the form given in the Theorem. We proceed now to derive that form. Let ω be any point satisfying

$$(\omega, p(t), 0) \in N^L_{G(t)}(x_*(t), f_t(x_*(t),u_*(t)),0).$$

According to Proposition 9.1, we have

$$(\omega, 0) \in \partial_L \big\{ \langle -p(t), f_t \rangle + K(t)|p(t)| d_{S(t)} \big\} (x_*(t), u_*(t)).$$

Since $(p'(t), 0)$ is a convex hull of such points $(\omega, 0)$, and since the convex hull of ∂_L is ∂_C, we obtain

$$(-p'(t), 0) \in \partial_C \big\{ \langle p(t), f_t \rangle - K(t)|p(t)| d_{S(t)} \big\} (x_*(t), u_*(t)) \tag{9.4}$$

(the minus sign can be brought out because of the calculus rule $\partial_C(-f) = -\partial_C f$). The Euler inclusion in the statement of the Theorem results from the sum rule for ∂_C together with the fact that $\partial_C d_{S(t)} \subset N^C_{S(t)}$.

There remains the final assertion of the Theorem, concerning the sequence of increasing radius functions. Applying the theorem for each R_i, we obtain a couple $(\lambda_{0,i}, p_i)$ satisfying (among other things) a Weierstrass condition of radius R_i as well as (9.3). Note that for each i, Prop. 9.1 applies with the data k^f_x, k^f_u, k_S corresponding to any given radius function, for example R_1 (say). Thus the uniform estimates used in the limiting procedure above are available, and the conclusion follows.

This completes the proof of the Theorem in the case $\Lambda \equiv 0$. However, the case in which a nonzero Λ is present is reducible to the already treated one by the familiar device of introducing an additional state coordinate y with dynamics $y'(t) = \Lambda_t(x(t), u(t))$, an initial condition $y(a) = 0$, and an additional term $y(b)$ in the cost (replacing the integral). Following the application of the Theorem (for this problem having zero integral cost), a straightforward decoding of notation leads to the full statement of Theorem 2.1. We omit these details.

Remark: The proof shows that the Euler inclusion can be stated in the potentially sharper form

$$(-p'(t), 0) \in \partial_C \big\{ \langle p(t), f_t \rangle - \lambda_0 \Lambda_t - K(t)|p(t)| d_{S(t)} \big\} (x_*(t), u_*(t)) \text{ a.e.}$$

Proof of Theorem 3.2. As in the proof of Theorem 2.1, we reduce to the case $\Lambda \equiv 0$. The following extra hypotheses allow us to adapt the proof of that theorem to the hybrid case.

Interim Hypotheses:

[IH1] For each t, the set $W(t)$ consists of finitely many points;
[IH2] There exists $C > 0$ such that, for almost all t:

$$\begin{gathered} w \in W(t) \cap B(w_*(t), R(t)),\ (x, v) \in S^{\epsilon, R}_*(t, w) \implies \\ |k^f_x(t, w_*(t)) - k^f_x(t, w)| + k_S(t)|k^f_v(t, w_*(t)) - k^f_v(t, w)| + |f_t(x, v, w) - x'_*(t)| \le C. \end{gathered}$$

We remark that the use of such interim hypotheses follows closely the lines of the proof of Theorem 5.1.2 in [5]. We modify the definition of the underlying multifunction F as follows, where $u = (v, w)$:

$$F(t, x) := \{ (f_t(x, u), \theta_t(u)) : (x, v) \in S(t, w),\ w \in W(t) \}.$$

Then $F(t, \cdot)$ continues to have locally closed graph, since $W(t)$ is a finite set (by [IH1]). The proof of Prop. 9.1 goes through without essential change and gives (9.1), for the modified function

$$k(t) := \sup_{w \in W(t)} k^f_x(t, w) + k_S(t) \sup_{w \in W(t)} k^f_v(t, w),$$

which is summable because of [IH2].

In the next step, the auxiliary problem is defined exactly as before. The necessary conditions again apply, and the main point is to derive the new hybrid adjoint inclusion from relation (9.3). This amounts to truncating its last component, the one involving w. The argument is the following.

Fixing a suitable t, let ω be any point satisfying (in partitioned notation)

$$(\omega, p(t), 0, 0) \in N^L_{G(t)}(x_*(t), f_t(x_*(t), u_*(t)), 0, 0).$$

Then $(\omega, p(t), 0, 0)$ is obtained as the limit of points

$$(\omega_i, p_i, a_i, b_i) \in N^P_{G(t)}(x_i, f_t(x_i, u_i), \theta_t(v_i, w_i)),$$

where

$$(x_i, v_i) \in S(t, w_i),\ (x_i, f_t(x_i, u_i), \theta_t(v_i, w_i)) \to (x_*(t), f_t(x_*(t), u_*(t)), 0, 0).$$

It follows that $v_i \to v_*(t)$ and that $w_i = w_*(t)$ for all i sufficiently large (since $W(t)$ is finite). Then (from the definition of proximal normal) we have

$$(\omega_i, p_i, a_i) \in N^P_{G_0(t)}(x_i, f_t(x_i, v_i, w_*(t)), c(t)(v_i - v_*(t))),$$

where $G_0(t)$ is the set

$$\{(x, f_t(x, v, w_*(t)), c(t)(v - v_*(t))) : (x, v) \in S(t, w_*(t))\}.$$

In the limit this gives the required 'truncated' inclusion:

$$(\omega, p(t), 0) \in N^L_{G_0(t)}(x_*(t), f_t(x_*(t), u_*(t)), 0),$$

which implies the hybrid adjoint inclusion, by the same arguments as in the previous proof.

In the last stage of the proof, we treat the problem without the Interim Hypotheses through inner approximation, as follows. It is possible to define finitely many measurable selections w_j of $W(\cdot)$ (one of which is $w_*(t)$) in such a way that the Interim Hypotheses hold when $W(t)$ is replaced by $\{w_j(t)\}$, and also such that the Weierstrass condition corresponding to the family $\{w_j(t)\}$ (when written for w_*) implies (to any specified tolerance) the one for $W(t)$. Since u_* remains optimal for this subproblem, the case of the theorem proven above can be invoked. There result multipliers which meet all the requirements, except that the Weierstrass condition holds only to the prescribed tolerance. The last step consists of a familiar sequential compactness argument. (See [5, pp. 207-209] for details.)

REFERENCES

[1] A. V. Arutyunov. *Optimality Conditions. Abnormal and Degenerate Problems.* Kluwer, Dordrecht, 2000.

[2] A. V. Arutyunov, D. Yu. Karamzin, and F. L. Pereira. Pontryagin's maximum principle for constrained impulsive control problems. Preprint 2009.

[3] F. H. Clarke. Maximum principles without differentiability. *Bulletin Amer. Math. Soc.*, 81:219–222, 1975.

[4] F. H. Clarke. The maximum principle under minimal hypotheses. *SIAM J. Control Optim.*, 14:1078–1091, 1976.

[5] F. H. Clarke. *Optimization and Nonsmooth Analysis.* Wiley-Interscience, New York, 1983. Republished as vol. 5 of *Classics in Applied Mathematics*, SIAM, 1990.

[6] F. H. Clarke. The maximum principle in optimal control. *Control and Cybernetics*, 34:709–722, 2005.

[7] F. H. Clarke. Necessary Conditions in Dynamic Optimization. *Memoirs of the Amer. Math. Soc.*, 173(816), 2005.

[8] F. H. Clarke and M. R. de Pinho. The nonsmooth maximum principle. *Control and Cybernetics.* In press.

[9] F. H. Clarke, Yu. Ledyaev, and M. R. de Pinho. An extension of the Schwarzkopf multiplier rule in optimal control. Preprint, 2009.

[10] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski. *Nonsmooth Analysis and Control Theory.* Graduate Texts in Mathematics, vol. 178. Springer-Verlag, New York, 1998.

[11] F. H. Clarke and R. B. Vinter. Optimal multiprocesses. *SIAM J. Control Optim.*, 27:1072–1091, 1989.

[12] M. R. de Pinho. Mixed constrained control problems. *J. Math. Anal. Appl.*, 278:293–307, 2003.

[13] M. R. de Pinho, M. M. A. Ferreira, and F. A. C. C. Fontes. Unmaximized inclusion necessary conditions for nonconvex constrained optimal control problems. *ESAIM Control Optim. Calc. Var.*, 11:614–632, 2005.

[14] M. R. de Pinho and A. Ilchmann. Weak maximum principle for optimal control problems with mixed constraints. *Nonlinear Anal.*, 48:1179–1196, 2002.

[15] M. R. de Pinho and J. F. Rosenblueth. Necessary conditions for constrained problems under Mangasarian-Fromowitz conditions. *SIAM J. Control Optim.*, 2009. In press.

[16] M. R. de Pinho and R. B. Vinter. An Euler-Lagrange inclusion for optimal control problems. *IEEE Trans. Automat. Control*, 40:1191–1198, 1995.

[17] M. R. de Pinho and R. B. Vinter. Necessary conditions for optimal control problems involving nonlinear differential algebraic equations. *J. Math. Anal. Appl.*, 212:493–516, 1997.

[18] M. R. de Pinho, R. B. Vinter, and H. Zheng. A maximum principle for optimal control problems with mixed constraints. *IMA J. Math. Control Inform.*, 18:189–205, 2001.

[19] E. N. Devdariani and Y. S. Ledyaev. Maximum principle for implicit control systems. *Appl. Math. Optim.*, 40:79–103, 1999.

[20] A. V. Dmitruk. Maximum principle for a general optimal control problem with state and regular mixed constraints. *Comp. Math. and Modeling*, 4:364–377, 1993.

[21] A. Ya. Dubovitskiĭ and A. A. Milyutin. Theory of the principle of the maximum. (Russian). In *Methods of the Theory of Extremal Problems in Economics.* Nauka, Moscow, 1981.

[22] R. V. Gamkrelidze. Optimal control processes with restricted phase coordinates. *Izv. Akad. Nauk SSSR, Ser. Mat.*, 24:315–356, 1960.

[23] M. R. Hestenes. *Calculus of Variations and Optimal Control Theory.* Wiley, New York, 1966.

[24] A. D. Ioffe and R. T. Rockafellar. The Euler and Weierstrass conditions for nonsmooth variational problems. *Calc. Var. Partial Differential Eq.*, 4:59–87, 1996.

[25] A. D. Ioffe and V. Tikhomirov. *Theory of Extremal Problems.* Nauka, Moscow, 1974. English translation, North-Holland (1979).

[26] K. Makowski and L. W. Neustadt. Optimal control problems with mixed control-phase variable equality and inequality constraints. *SIAM J. Control*, 12:184–228, 1974.

[27] A. A. Milyutin and N. P. Osmolovskii. *Calculus of Variations and Optimal Control.* American Math. Soc., Providence, 1998.

[28] L. W. Neustadt. *Optimization.* Princeton University Press, Princeton, 1976.

[29] Z. Páles and V. Zeidan. First- and second-order necessary conditions for control problems with constraints. *Trans. Amer. Math. Soc.*, 346:421–453, 1994.

[30] Z. Páles and V. Zeidan. Strong local optimality conditions for control problems with mixed state-control constraints. In *Proc. Conf. Decision and Control, Las Vegas*, pages 4738–4743. IEEE, 2002.

[31] Z. Páles and V. Zeidan. Optimal control problems with set-valued control and state constraints. *SIAM J. Optim.*, 14:334–358, 2003.

[32] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko. *The Mathematical Theory of Optimal Processes.* Wiley-Interscience, New York, 1962.

[33] A. B. Schwarzkopf. Optimal controls with equality state constraints. *J. Optim. Th. Appl.*, 19:455–468, 1976.

[34] G. Stefani and P. Zezza. Optimality conditions for a constrained control problem. *SIAM J. Control Optim.*, 34:635–659, 1996.

[35] R. B. Vinter. *Optimal Control.* Birkhäuser, Boston, 2000.