AN EXTENSION OF THE SCHWARZKOPF MULTIPLIER RULE IN OPTIMAL CONTROL

FRANCIS CLARKE*, YURI LEDYAEV†, AND MARIA DO ROSÁRIO DE PINHO‡

Abstract. The search for multiplier rules in dynamic optimization has been an important theme in the subject for over a century; it was a central one in the classical calculus of variations, and the Pontryagin maximum principle of optimal control theory is part of this quest. A more recent thread has involved problems with so-called mixed constraints involving the control and state variables jointly, a subject which now boasts a considerable literature. Recently, Clarke and de Pinho proved a general multiplier rule for such problems that extends and subsumes rather directly most of the available results, namely those which postulate some kind of rank condition or, more generally, a constraint qualification (or generalized Mangasarian-Fromowitz condition). An exception to this approach is due to Schwarzkopf, whose well-known theorem replaces the rank hypothesis, for relaxed problems, by one of covering. The purpose of this article is to show how to obtain this type of theorem from the general multiplier rule of Clarke and de Pinho. In so doing, we subsume, extend and correct the currently available versions of Schwarzkopf's result.

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

AMS subject classifications. 49K15, 49K21

1. Introduction.

A long-standing problem in the calculus of variations has been to establish strong multiplier rules for problems in which constraints such as $\phi(t, x(t), x'(t)) = 0$ are imposed along the competing arcs $x(\cdot)$. Such multiplier rules have been obtained under *rank hypotheses*, requiring for example that the Jacobian matrix $D_v\phi(t, x, v)$ have maximal rank in a certain region. In the setting of a standard optimal control problem, the so-called *mixed constraints* involve such conditions as $\phi(t, x(t), u(t)) = 0$, where u is the control variable. Again, rank hypotheses are typically made in order to derive multiplier rules in a context related to the Pontryagin maximum principle. When the constraints are described in more general fashion, for example by a combination of equalities, inequalities, and unilateral inclusions, the rank hypotheses are replaced by appropriate constraint qualifications, frequently of Mangasarian-Fromowitz type. A considerable literature now exists on this topic: see [1, 2, 3, 4, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 24].

Recently, Clarke and de Pinho [5] proved a general multiplier rule for such problems that extends and subsumes rather directly the majority of known results, which postulate some kind of rank condition or constraint qualification. An exception to this approach is due to Schwarzkopf [22, 23] (see also [11]), whose well-known theorem replaces the rank hypothesis, for relaxed problems, by one of covering. The purpose of this article is to show how to obtain this type of theorem from the general multiplier rule of Clarke and de Pinho. In so doing, we subsume, extend and correct the currently available versions of Schwarzkopf's result.

We shall make use below of certain constructs of nonsmooth analysis, in particular the generalized (Clarke) normal cone $N_S^C(x)$ and the limiting normal cone $N_S^L(x)$ to a closed set S at a point $x \in S$, as well as the generalized gradient $\partial_C f(x)$ and the limiting subdifferential $\partial_L f(x)$. These concepts are recalled briefly in [5] and presented

*Institut universitaire de France and Université de Lyon, France (clarke@math.univ-lyon1.fr).

†Department of Mathematics, Western Michigan University, USA (iouri.lediaev@wmich.edu).
Partially supported by National Science Foundation Grant DMS-0708019.

‡Universidade do Porto, Faculdade de Engenharia, Oporto, Portugal (mrpinho@fe.up.pt). Partially supported by FCT Project PTDC/EEACRO/116014/2009.

in detail in [6]; we mention that they all reduce to classical notions in the presence of smoothness or convexity. For this reason, our theorem statements do not require familiarity with nonsmooth analysis in order to be understood.

We denote by $B(x, r)$ the closed ball of radius r centered at x. The notation $|\cdot|$ always refers to the Euclidean norm.

2. Preliminaries. We recall here the main theorem of [5].[1] 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

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

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

$$(P) \quad \begin{cases} \text{Minimize } \ell(x(a), x(b)) \\ \text{subject to} \\ \qquad\quad x'(t) \;=\; f_t(x(t), u(t)) \quad \text{a.e.} \quad t \in [a, b] \\ \ (x(t), u(t)) \;\in\; S(t) \qquad\qquad\quad\ \text{a.e.} \quad t \in [a, b] \\ \ (x(a), x(b)) \;\in\; E. \end{cases}$$

Notice that the t-dependence of 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 are the following: f is $\mathcal{L}\times\mathcal{B}$ measurable[2]; 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. The theorems below feature hypotheses directly related to a given pair (x_*, u_*) that is admissible for (P).

Let $R > 0$. We say that (x_*, u_*) is a *local minimum of radius* R for (P) provided that for some $\epsilon > 0$, for every pair (x, u) admissible for (P) which also satisfies

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

we have $\ell(x(a), x(b)) \ge \ell(x_*(a), x_*(b))$. The hypothesis that (x_*, u_*) is a local minimum in this sense is strictly weaker than the more familiar ones (strong or $W^{1,1}$).

We define

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

We assume that $S(t)$ is locally closed at each point $(x, u) \in S_*^{\epsilon,R}(t)$.

The two main hypotheses of the theorem concern *Lipschitz behavior* of f with respect to (x, u) and a certain *bounded slope condition* bearing upon the sets $S(t)$.

[1] We omit the integral cost term, however, since it is not needed here, and we further specialize to the case of a finite constant radius R rather than a time-dependent one.

[2] 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 $\mathbb{R}^n\times\mathbb{R}^m$.

$[\mathbf{L}_*^{\epsilon,\mathbf{R}}]$: There exist summable functions k_x^f, k_u^f 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

$$|f_t(x_1,u_1)-f_t(x_2,u_2)| \le k_x^f(t)|x_1-x_2| + k_u^f(t)|u_1-u_2|.$$

Concerning the mixed constraint, the hypothesis is the following:

$[\mathbf{BS}_*^{\epsilon,\mathbf{R}}]$: There exists a number k_S such that, for almost every t, the following bounded slope condition holds:

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

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 [4] as *stratified.* This feature plays an important role in what is to come.

THEOREM 2.1. *Let (x_*,u_*) be a local minimum of radius R for (P), where, for some $\epsilon>0$, $[\mathrm{BS}_*^{\epsilon,R}]$ holds and f satisfies $[\mathrm{L}_*^{\epsilon,R}]$. Then there exist an arc p and a nonnegative number λ_0 satisfying the* **nontriviality condition**

$$\lambda_0+\|p\|_\infty = 1,$$

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:*

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

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

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

A special case. When the mixed constraint $(x,u)\in S(t)$ has special structure defined through functional equalities and inequalities, 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 appearing in the Euler inclusion via multipliers. This theme is developed in detail in [5]. It turns out that various types of results encountered in the literature are subsumed by this approach. However, we require for the purposes of this article a special case concerning the following problem that features affine structure relative to the control variable $c=(c^0,c^1,\dots,c^N)\in\mathbb{R}^{N+1}$: to minimize $\ell(x(a),x(b))$ subject to $(x(a),x(b))\in E$ and

$$x'(t) = f_t(x(t),c(t)) := \sum_i c^i g_t^i(x(t)) \text{ a.e.}$$

$$\phi_t(x(t),c(t)) := \sum_i c^i\theta_t^i(x(t)) \in \Phi_t,\ \ c(t)\in\Sigma \text{ a.e.}$$

Here, $g^i:[a,b]\times\mathbb{R}^n\to\mathbb{R}^n$ and $\theta^i:[a,b]\times\mathbb{R}^n\to\mathbb{R}^\kappa$ $(i=0,1,\dots,N)$ are given functions, Φ is a multifunction from $[a,b]$ to the closed subsets of $\mathbb{R}^\kappa$ (all the data is taken $\mathcal{L} \times \mathcal{B}$ measurable), and $\Sigma \subset \mathbb{R}^{N+1}$ is a given compact convex set.[3] The basic hypotheses remain in force and, as before, we let (x_*, c_*) be a local minimum of radius R for the problem. The problem is clearly the special case of the general problem (P) in which (identifying c with u)

$$S(t) := \{(x, u) : u \in \Sigma,\ \phi_t(x, u) \in \Phi_t\}$$

(and in which f has the indicated structure).

We assume that for a certain summable function k^g and constant k^θ, for almost every t in $[a, b]$, for each $i = 0, 1 \ldots, N$, we have, for all x, y in a neighborhood of $B(x_*(t), \epsilon)$:

$$|g_t^i(x) - g_t^i(y)| \ \le\ k_t^g |x - y|,\ \ |\theta_t^i(x) - \theta_t^i(y)| \ \le\ k^\theta |x - y|.$$

It is assumed in addition that for each i, the function $x \mapsto \theta_t^i(x)$ is *strictly differentiable* at $x_*(t)$ a.e., and that the function $t \mapsto g_t^i(x_*(t))$ is summable. We recall that a function $\psi : \mathbb{R}^n \to \mathbb{R}^k$ is strictly differentiable at x if the derivative $\psi'(x)$ exists and

$$\lim_{y,\, z \to x,\, y \neq z} |\psi(z) - \psi(y) - \psi'(x)(z - y)|/|z - y| = 0.$$

This is a stricter requirement than mere differentiability, and a weaker one than continuous differentiability.

We posit the following *calibrated* constraint qualification of the type introduced in [5]: for some constant M, for almost every t,

$$\begin{aligned} &(x, c) \in S_*^{\epsilon, R}(t),\ \lambda \in N_{\Phi_t}^C(\phi_t(x, c)),\ \gamma \in N_\Sigma^C(c), \\ &\qquad \beta \ =\ \nabla_c\{\langle \lambda, \phi_t \rangle\}(x, c) + \gamma \ =\ (\lambda \cdot \theta_t^0(x), \lambda \cdot \theta_t^1(x), \ldots, \lambda \cdot \theta_t^N(x)) + \gamma \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \implies |\lambda| \le M|\beta|. \end{aligned} \tag{2.1}$$

The following result paves the way for the next section, but is of independent interest. Note that, in contrast to Theorem 2.1, the Euler inclusion below has been projected to involve only its first component. Further, the Weierstrass inequality, which involves control values c for which $\phi_t(x_*(t), c)$ may not belong to Φ_t (thus, inadmissible ones), is asserted globally, and not just to radius R. It is the special structure of the problem that makes these modifications possible.

COROLLARY 2.2. *Under the hypotheses above, there exist an arc p and a nonnegative number λ_0 satisfying the nontriviality and transversality conditions of Theorem 2.1, as well as a summable function $\lambda : [a, b] \to \mathbb{R}^\kappa$ satisfying*

$$\lambda(t) \in N_{\Phi_t}^C(\phi_t(x_*(t), c_*(t))) \text{ a.e.}$$

such that the adjoint inclusion takes the explicit multiplier form

$$-p'(t) \in \partial_C\{\ \langle p(t), f_t(\cdot, c_*(t)) \rangle - \langle \lambda(t), \phi_t(\cdot, c_*(t)) \rangle\ \}(x_*(t)) \text{ a.e.}$$

and such that, for almost every t, the following extended Weierstrass condition holds:

$$\begin{aligned} c \in \Sigma \implies &\langle p(t), f_t(x_*(t), c) \rangle - \langle \lambda(t), \phi_t(x_*(t), c) \rangle \\ &\qquad \le \langle p(t), f_t(x_*(t), c_*(t)) \rangle - \langle \lambda(t), \phi_t(x_*(t), c_*(t)) \rangle\,. \end{aligned}$$

[3] The control set Σ is allowed to be of the form $\{1\} \times B(0, 1)$, for example, forcing $c_0 \equiv 1$. Thus the problem may incorporate a 'drift term'.

Proof. We shall justify applying Theorem 2.1. Observe first that our hypotheses imply that the set $S(t)$ is locally closed at each $(x,c) \in S_*^{\epsilon,R}(t)$. Now let σ satisfy $c \in \Sigma \Longrightarrow |c| \le \sigma$, and set

$$m(t) := \sum_{i=0}^{N} \max_{x \in B(x_*(t),\epsilon)} |g_t^i(x)| \in L^1(a,b).$$

It is evident that f satisfies [$\mathrm{L}_*^{\epsilon,R}$] with $k_x^f = \sigma(N+1)k^g$ and $k_u^f = m$, so we turn to [$\mathrm{BS}_*^{\epsilon,R}$]. Let t be such that the Lipschitz condition and (2.1) hold. Let $(x,c) \in S_*^{\epsilon,R}(t)$ and $(\alpha,\beta) \in N_{S(t)}^P(x,c)$. Then, by known results in nonsmooth analysis,[4] there exists

$$(\lambda,\gamma) \in N_{\Phi_t}^L(\phi_t(x,c)) \times N_\Sigma^C(c)$$

such that

$$(\alpha,\beta) \in \partial_C \left\{ \langle \lambda, \phi_t \rangle + \langle \gamma, c \rangle \right\}(x,c). \tag{2.2}$$

If a function $h(x,c)$ of two variables is of the form $\langle F(x), c \rangle$ and $(\alpha,\beta) \in \partial_C h(x,c)$ (jointly), then $\alpha \in \partial_C \langle F(x), c \rangle$ (taken in x) and $\beta = F(x)$. (This follows, for example, from the Gradient Formula for generalized gradients [6, Th. 2.8.1].) In view of this fact, looking at the second component of (2.2), we derive

$$\beta \ = \ (\lambda \cdot \theta_t^0(x), \lambda \cdot \theta_t^1(x), \ldots, \lambda \cdot \theta_t^N(x)) + \gamma$$

which, in view of (2.1), yields $|\lambda| \le M|\beta|$. Observe next that the first component implies

$$\alpha \in \partial_C \left\{ \langle \lambda, \phi_t(\cdot, c) \rangle \right\}(x),$$

which yields the bound $|\alpha| \le \sigma k^\theta (N+1)|\lambda|$ (by estimating the Lipschitz constant). Combining the last two inequalities, we discover [$\mathrm{BS}_*^{\epsilon,R}$], with $k_S := \sigma k^\theta (N+1) M$.

The hypotheses of Theorem 2.1 are all satisfied, and we deduce the existence of an arc p and a number $\lambda_0 \ge 0$ satisfying the nontriviality and transversality conditions of Theorem 2.1, as well as the Euler and Weierstrass conditions. We know as above that for almost every t, for any element (α,β) of $N_{S(t)}^C(x_*(t), c_*(t))$, there exists $(\lambda(t),\gamma(t))$ such that

$$(\lambda(t),\gamma(t)) \in N_{\Phi_t}^C(\phi_t(x_*(t),c_*(t))) \times N_\Sigma^C(c_*(t))$$

and such that

$$\begin{aligned}(-p'(t),\gamma) &\in \partial_C \{\langle p(t), f_t \rangle\}(x_*(t),c_*(t)) - \nabla_{x,c}\{\langle \lambda(t), \phi_t \rangle\}(x_*(t),c_*(t)) \\ &= \partial_C \{\langle p(t), f_t \rangle - \langle \lambda(t), \phi_t \rangle\}(x_*(t),c_*(t)) \text{ a.e.}\end{aligned}$$

(We have used here the fact that ϕ is strictly differentiable at $(x_*(t), c_*(t))$ a.e.) The functions involved may be taken to be measurable. Looking at the first component in this inclusion gives the required adjoint equation, and, upon writing for γ the definition of normal vector to the convex set Σ, the second component gives precisely

[4] For example, one may apply Prop. 4.1 of [5] with $u = (x,c)$, $\phi(u) = (\phi(x,c),c)$, $\Phi = \Phi_t \times \Sigma$.

the stated Weierstrass condition. The second component of the inclusion also shows that (for almost every t) there exists $\zeta(t)$ satisfying

$$|\zeta(t)| \le |p(t)|m(t),\ \zeta(t) \ = \ \nabla_c\{\langle\lambda(t), \phi_t\rangle\}(x_*(t), c_*(t)) + \gamma.$$

It follows from (2.1) that $|\lambda(t)| \le M|p(t)|m(t)$, so that λ is summable. □

Remark. As shown in [5], the constraint qualification (2.1) follows from certain sets of additional (and more easily verifiable) hypotheses on the data. For example, let the θ^i be continuously differentiable and $\Phi_t \equiv \{0\}$. Then (2.1) holds (by [5, Prop. 4.4]) if there exist positive r and δ such that, for every t and for every $x \in B(x_*(t), \epsilon)$, we have

$$\left\{\sum_i c^i\theta_t^i(x) : |c| \le \delta\right\} \ \supset \ B(0, r).$$

This type of covering condition is precisely what is substantially generalized in the next section.

3. The main theorem. We consider now the following optimal control problem:

$$(Q) \quad \left\{\begin{array}{rcll} \multicolumn{4}{l}{\text{Minimize } \ell(x(a), x(b))} \\ \multicolumn{4}{l}{\text{subject to}} \\ x'(t) & = & f_t(x(t), u(t)) & \text{a.e.} \quad t \in [a, b] \\ \theta_t(x(t), u(t)) & \in & \Phi_t & \text{a.e.} \quad t \in [a, b] \\ u(t) & \in & U_t & \text{a.e.} \quad t \in [a, b] \\ (x(a), x(b)) & \in & E. & \end{array}\right.$$

Here, $f : [a, b] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\theta : [a, b] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^\kappa$ are given functions, Φ is a multifunction from $[a, b]$ to the subsets of $\mathbb{R}^\kappa$, and U is a multifunction from $[a, b]$ to the subsets of $\mathbb{R}^m$. We assume that the admissible process (x_*, u_*) is a local minimum in the following sense: for some $\epsilon > 0$, for every pair (x, u) admissible for (Q) which also satisfies

$$x(t) \in B(x_*(t), \epsilon) \ \ \forall\, t, \qquad \int_a^b |x'(t) - x'_*(t)|\, dt \le \epsilon,$$

we have $\ell(x(a), x(b)) \ge \ell(x_*(a), x_*(b))$.

Hypotheses for Theorem 3.1.

[H1] The function ℓ is locally Lipschitz, and the set E is closed.

[H2] The function $(t, u) \mapsto (f_t(x, u), \theta_t(x, u))$ is $\mathcal{L} \times \mathcal{B}$ measurable for each x; the multifunction Φ is measurable and closed-valued; the graph of U is $\mathcal{L} \times \mathcal{B}$ measurable.

[H3] There exists a summable function k^f such that, for almost every t, for every $u \in U_t$, the function $f_t(\cdot, u)$ is Lipschitz with constant k_t^f on $B(x_*(t), \epsilon)$.

[H4] There exists a constant k^θ such that, for almost every t, for every $u \in U_t$, the function $\theta_t(\cdot, u)$ is Lipschitz with constant k^θ on $B(x_*(t), \epsilon)$, and this function is strictly differentiable at $x_*(t)$.

[H5] There exist a summable function r and $\delta > 0$ such that the multifunction $U_r(t) := \{u \in U_t : |f_t(x_*(t), u)| \le r(t)\}$ satisfies

$$\theta_t(x_*(t), U_r(t)) \supset B(\theta_t(x_*(t), u_*(t)), \delta) \text{ a.e.}$$

[H6] For almost every t, for every $x \in B(x_*(t), \epsilon)$, the following set is convex:

$$\{(f_t(x,u), \theta_t(x,u)) : u \in U_t\}.$$

THEOREM 3.1. *Under the hypotheses above, there exist an arc p and a nonnegative number λ_0 satisfying the nontriviality and transversality conditions of Theorem 2.1, as well as a summable function $\lambda : [a,b] \to \mathbb{R}^\kappa$ satisfying*

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

such that the adjoint inclusion takes the explicit multiplier form

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

and such that, for almost every t, the following extended Weierstrass condition holds:

$$\begin{aligned} u \in U_t \implies & \langle p(t), f_t(x_*(t), u)\rangle - \langle \lambda(t), \theta_t(x_*(t), u)\rangle \\ & \leq \langle p(t), f_t(x_*(t), u_*(t))\rangle - \langle \lambda(t), \theta_t(x_*(t), u_*(t))\rangle . \end{aligned} \quad (3.1)$$

Proof. There is no loss of generality in assuming that $\theta_t(x_*(t), u_*(t)) = 0$ a.e., since we can simply redefine θ and Φ as follows:

$$\tilde{\theta}_t(x,u) := \theta_t(x,u) - \theta_t(x_*(t), u_*(t)),\ \tilde{\Phi}_t := \Phi_t - \theta_t(x_*(t), u_*(t)),$$

and both the hypotheses and conclusions are robust with respect to this modification. We may also suppose that the Lipschitz property in [H3][H4] holds on a neighborhood of $B(x_*(t), \epsilon)$.

The canonical basis vectors of $\mathbb{R}^\kappa$ are denoted $\{e_1, e_2, \ldots, e_\kappa\}$. According to [H5], and by measurable selections, there exist control functions $u_1^+(t)$ and $u_1^-(t)$ with values in U_r such that, for almost every t, we have $\theta_t(x_*(t), u_1^+(t)) = \delta e_1$ and $\theta_t(x_*(t), u_1^-(t)) = -\delta e_1$. We choose analogously two control functions u_i^+ and u_i^- for each canonical vector e_i.

In light of [H4] it follows that, by suitably shrinking ϵ if necessary, we can arrange that for almost every t we have, for each $i = 1, 2 \ldots, \kappa$:

$$|\theta_t(x, u_i^+(t)) - \delta e_i| < \delta/(6\kappa),\ |\theta_t(x, u_i^-(t)) + \delta e_i| < \delta/(6\kappa) \ \ \forall\, x \in B(x_*(t), \epsilon). \quad (3.2)$$

We proceed now to define another set of control functions figuring in the auxiliary problem being constructed. Let $d > 0$ (later d will decrease to 0), and define

$$\begin{aligned} f_t^* &:= f_t(x_*(t), u_*(t)) \\ h_t(p, \lambda, u) &:= \langle p, f_t(x_*(t), u) - f_t^*\rangle - \langle \lambda, \theta_t(x_*(t), u)\rangle \\ U_t^d &:= \{u \in U_t : |(f_t(x_*(t), u) - f_t^*, \theta_t(x_*(t), u))| \leq 1/d\}, \\ H_t^d(p, \lambda) &:= \sup \big\{ h_t(p, \lambda, u) : u \in U_t^d \big\}. \end{aligned}$$

Let $\{(p_j, \lambda_j)\}$ be a finite subset of $B(0,1) \subset \mathbb{R}^n \times \mathbb{R}^\kappa$ such that the union of the balls $B((p_j, \lambda_j), d^2)$ covers $B(0,1)$. We choose to label these points with indices j ranging from $2\kappa + 1$ to a certain integer N_d. Measurable selection theory implies the existence of control functions v_j, $j = 2\kappa + 1, 2\kappa + 2, \ldots, N_d$ with values in U_t^d having the property that for almost every t we have

$$h_t(p_j, \lambda_j, v_j(t)) > H_t^d(p_j, \lambda_j) - d. \quad (3.3)$$

We now proceed to define a problem having the affine structure that will allow us to invoke Corollary 2.2. The functions of the problem are given by

$$\begin{array}{llll} g_t^0(x) & := f_t(x, u_*(t)), & \theta_t^0(x) := \theta_t(x, u_*(t)) \\ g_t^i(x) & := f_t(x, u_i^+(t)) - f_t^*, & \theta_t^i(x) := \theta_t(x, u_i^+(t)),\ i = 1, 3, \ldots, 2\kappa - 1 \\ g_t^i(x) & := f_t(x, u_i^-(t)) - f_t^*, & \theta_t^i(x) := \theta_t(x, u_i^-(t)),\ i = 2, 4, \ldots, 2\kappa \\ g_t^i(x) & := f_t(x, v_i(t)) - f_t^*, & \theta_t^i(x) := \theta_t(x, v_i(t)),\ i = 2\kappa + 1, 2\kappa + 2, \ldots, N_d, \end{array}$$

and we define the control set by

$$\Sigma := \left\{ c = (c_0, c_1, \ldots, c_{N_d}) : c_0 = 1, c_i \geq 0\ (i = 1, 2, \ldots, N_d),\ \sum_1^{N_d} c_i \leq 1 \right\}.$$

The cost function and boundary constraints of this new problem are those of the original problem (Q); Φ is also unchanged. It follows from [H6] that any admissible state trajectory x for this problem is an admissible state trajectory for (Q), from which we deduce that a (local) solution to the problem is obtained by taking the control $c_*(t) := (1, 0, \ldots, 0)$ and its corresponding state trajectory x_*. Evidently, for any $R > 0$, the process (x_*, c_*) is a local minimum of radius R. We shall exploit this fact presently in applying Corollary 2.2; the point is that this will require verifying the constraint qualification (2.1) only for points c near c_*. Specifically, we choose R small enough so that

$$c \in \Sigma,\ |c - c_*| \leq R \implies \max_{1 \leq i \leq N_d} c_i < 1,\ \sum_1^{N_d} c_i < 1.$$

We observe that the normal cone at any such point c to the convex set Σ defined above consists of vectors of the form $(\eta, \gamma) \in \mathbb{R} \times \mathbb{R}^{N_d}$ for which $\gamma \leq 0$ (in the vector sense).

We proceed now to verify the hypothesis (2.1) of Corollary 2.2. Let (x, c), λ, and β be as described there. With the above in mind, and looking for the moment at only the second and third coordinates of the inclusion in (2.1) (for indices 1 and 2), we find (in light of (3.2)):

$$\beta_1 = \delta\lambda_1 + \mu_1|\lambda| + \gamma_1,\ \beta_2 = -\delta\lambda_1 + \mu_2|\lambda| + \gamma_2, \tag{3.4}$$

where $|\mu_1|, |\mu_2| < \delta/(6\kappa)$ and $\gamma_1, \gamma_2 \leq 0$. These equations imply $\gamma_1 + \gamma_2 = \beta_1 + \beta_2 - (\mu_1 + \mu_2)|\lambda|$, whence

$$|\gamma_1| \leq |\beta_1 + \beta_2| + |\mu_1 + \mu_2||\lambda| \leq 2|\beta| + \delta|\lambda|/(3\kappa).$$

Substituting this in the first of the preceding equalities, we obtain $|\lambda_1| \leq 3|\beta|/\delta + |\lambda|/(2\kappa)$. A similar argument holds for the other coordinates of λ, and by summing we obtain $|\lambda| \leq 3\kappa|\beta|/\delta + |\lambda|/2$. This confirms that the constraint qualification (2.1) holds, with $M := 6\kappa/\delta$.

The other hypotheses of Corollary 2.2 are easily verified (the summability of the functions $|f_t(x_*(t), u_i^+(t))|$, $|f_t(x_*(t), u_i^-(t))|$ and f_t^* is used here), so we deduce the existence of p, λ_0 and λ as described there. It is easy to see that the adjoint equation is precisely the one stated in the theorem. The Weierstrass condition implies (for almost every t) the following first-order necessary condition:

$$\begin{aligned}\beta &:= \Big(\big\langle p(t), g_t^1(x_*(t))\big\rangle, \big\langle p(t), g_t^2(x_*(t))\big\rangle, \ldots, \Big\langle p(t), g_t^{N_d}(x_*(t))\Big\rangle\Big)\\ &= \big(\big\langle p(t), f_t(x_*(t), u_1^+(t))\big\rangle, \big\langle p(t), f_t(x_*(t), u_1^-(t))\big\rangle, \ldots, \langle p(t), f_t(x_*(t), v_{N_d}(t))\rangle\big)\\ &= \Big(\big\langle \lambda(t), \theta_t^1(x_*(t))\big\rangle, \big\langle \lambda(t), \theta_t^2(x_*(t))\big\rangle, \ldots, \Big\langle \lambda(t), \theta_t^{N_d}(x_*(t))\Big\rangle\Big) + \gamma(t),\end{aligned}$$

where $\gamma(t) \le 0$. As shown above, this implies

$$|\lambda(t)| \le 6\kappa|(\beta_1, \beta_2, \ldots, \beta_{2\kappa})|/\delta \;\le\; (12\kappa^2/\delta)r(t), \tag{3.5}$$

since the u_i^+, u_i^- have values in U_r and $|p(t)| \le 1$.

For almost every t, the Weierstrass condition also yields (by taking

$$c = (1, 0, \ldots, 1, \ldots, 0),$$

where the second 1 is in the jth position):

$$\begin{aligned}h_t(p(t), \lambda(t), v_j(t)) = \langle p(t), f_t(x_*(t), v_j(t)) - f_t^*\rangle - \langle \lambda(t), \theta_t(x_*(t), v_j(t))\rangle\\ \le 0 = h_t(p(t), \lambda(t), u_*(t)),\; j = 2\kappa+1, 2\kappa+2, \ldots, N_d.\end{aligned}$$

Set

$$\Lambda_t := (12\kappa^2/\delta)r(t) + 1,$$

so that $|(p(t), \lambda(t))| \le \Lambda_t$ a.e. by (3.5). Fix a value of t for which all estimates hold. There exists an index j for which $|(p(t), \lambda(t)) - \Lambda_t(p_j, \lambda_j)| < d^2\Lambda_t$. Then

$$\begin{aligned}0 \;&\ge\; h_t(p(t), \lambda(t), v_j(t)) \;\ge\; \Lambda_t h_t(p_j, \lambda_j, v_j(t)) - d^2\Lambda_t/d\\ &\ge\; \Lambda_t H_t^d(p_j, \lambda_j) - 2d\Lambda_t \text{ (see (3.3))}\\ &= H_t^d(\Lambda_t(p_j, \lambda_j)) - 2d\Lambda_t \;\ge\; H_t^d(p(t), \lambda(t)) - 3d\Lambda_t,\end{aligned}$$

in view of how U_t^d is defined, and since $|(p(t), \lambda(t)) - \Lambda_t(p_j, \lambda_j)| < d^2\Lambda_t$.

Summarizing, we have p, λ_0 and λ satisfying:

$$\|p\|_\infty + \lambda_0 = 1 \tag{3.6}$$

$$(p(a), -p(b)) \;\in\; \partial_L \lambda_0 \ell(x_*(a), x_*(b)) + N_E^L(x_*(a), x_*(b)) \tag{3.7}$$

$$-p'(t) \in \partial_C\big\{\, \langle p(t), f_t(\cdot, u_*(t))\rangle - \langle \lambda(t), \theta_t(\cdot, u_*(t))\rangle \,\big\}(x_*(t)) \text{ a.e.} \tag{3.8}$$

$$H_t^d(p(t), \lambda(t)) \le 3d\Lambda_t \text{ a.e.} \tag{3.9}$$

These would be exactly the required conclusions if d were 0. To conclude the proof, we now allow d to decrease to 0 along a sequence d_i. In view of the adjoint equation, the uniform L^1 estimate (3.5), and the Dunford-Pettis criterion for weak compactness, we may assume by taking subsequences that the resulting sequence p_i converges uniformly to an arc p, λ_i converges weakly in L^1 to λ, and λ_{0_i} to λ_0; it follows from standard arguments that (3.6) to (3.9) hold in the limit with $d = 0$.

□

Remark on [H5]. If we assume additionally that all the data are continuous and that U is autonomous and compact, then [H5] is evidently equivalent to the following:

[H5]$'$ For some $\delta > 0$, we have

$$\theta_t(x_*(t), U) \supset B(\theta_t(x_*(t), u_*(t)), \delta) \text{ a.e.}$$

(that is, we may take $U_r = U$ in [H5]). This is the type of covering hypothesis made in Schwarzkopf's article [22], in the more restrictive setting of a single equality constraint. In the absence of such additional hypotheses, however, and other things being equal, it turns out that hypothesis [H5]$'$ is strictly weaker than [H5]. In fact, Theorem 3.1 *fails* if [H5] is replaced by [H5]$'$, contrary to the assertion in [11]; we now provide a counterexample.

We take the state and the control in one dimension, $[a, b] = [0, 1]$, with

$$x' = f_t(x, u) = x + u,\ \phi_t(x, u) = tu,\ U_t = \mathbb{R}.$$

We seek to minimize $-x(1)$ subject to $x(0) = 0$, $\phi_t(x, u) = 0$ a.e. (so that $\Phi \equiv \{0\}$). It is clear that the only admissible process is $x_* = u_* = 0$, which is thereby optimal.

All the hypotheses of Theorem 3.1 are evidently satisfied, except that [H5]$'$ holds (for arbitrary positive δ) rather than [H5]. Let (p, λ_0, λ) satisfy the conclusions of that theorem. Then the adjoint equation and transversality yield $-p' = p$, $p(1) = \lambda_0$, so that $p(t) = \lambda_0 e^{1-t}$. From nontriviality we deduce $\lambda_0 > 0$. The Weierstrass condition (3.1) implies $p(t) - \lambda(t)t = 0$ a.e., whence $\lambda(t) = \lambda_0 e^{1-t}/t$ a.e. But this function fails to lie in L^1: the conclusions of the theorem fail to hold.

Remark on [H6]. As shown by an example in [11], Theorem 3.1 may fail in the absence of the convexity hypothesis [H6]. However, it has been an open question as to whether the theorem might be true without [H6] if the Weierstrass condition (3.1) is replaced by the following weaker (nonextended) one:

$$u \in U_t,\ \theta_t(x_*(t), u) \in \Phi_t \implies \langle p(t), f_t(x_*(t), u)\rangle \ \le\ \langle p(t), f_t(x_*(t), u_*(t))\rangle\,.$$

(This is implied by (3.1) whenever Φ has convex values, as for example in the case of equality and inequality constraints.) We now show by a counterexample that the answer to the question is negative.

We take the state in one dimension, $[a, b] = [0, 1]$, and the control (u, v) in two dimensions. We set:

$$x' = f_t(x, u, v) = u,\ \theta_t(u, v) = u(x + v^2),\ U = \text{unit ball in } \mathbb{R}^2,\ \Phi \equiv \{0\},$$

and we seek to minimize $-x(1)$ subject to $x(0) = 0$, $\theta_t(x, u, v) = 0$ a.e. We observe that all admissible states x remain nonpositive, for if x is zero at c and positive on (c, d), then $x + v^2 > 0$ on (c, d), forcing $u = 0$ there; but then $x \equiv 0$ on $[\,c, d]$, contradiction. It follows that the state $x_* = 0$ and the control $u_* = v_* = 0$ are optimal.

All the hypotheses of Theorem 3.1 are evidently satisfied, except [H6]. (Note that [H5] holds with $U_r = U$.) Let p, λ_0, λ satisfy the conclusions of that theorem. Then we have $p \equiv \lambda_0$, so p is a positive constant. The maximization in the weaker Weierstrass condition implies $p = 0$, contradiction.

Tangential covering conditions. On the basis of known results on restricted implicit function theorems[5], it is natural to suspect that the covering condition [H5]

[5] See for example [6, Theorem 3.3.4].

might be replaced with a more subtle one involving normals or tangents to the set Φ_t, for example:

$$\left\{\theta_t(x_*(t),u) - T^C_{\Phi_t}(\theta_t(x_*(t),u)) : u \in U_r(t)\right\} \supset B(\theta_t(x_*(t),u_*(t)),\delta) \text{ a.e.},$$

where T^C denotes the (generalized) tangent cone polar to N^C. This is evidently less restrictive than [H5], since the tangent cone contains 0. However, extra hypotheses on Φ must then be added to ensure that this leads to the required local (and uniform) constraint qualification when $x_*(t)$ is replaced by nearby values of x.

One approach leading to an easily-stated result is to postulate that Φ_t is a closed convex cone for each t (which covers in particular the case when the mixed constraints are given by functional equalities and inequalities). We then have $T^C_{\Phi_t}(z) \supset \Phi_t$ at all points $z \in \Phi_t$, and the following hypothesis suggests itself:

[H5]$_{\mathbf{T}}$ There exist a summable function r and $\delta > 0$ such that the multifunction $U_r(t) := \{u \in U_t : |f_t(x_*(t),u)| \le r(t)\}$ satisfies

$$\left\{\theta_t(x_*(t),u) - \Phi_t : u \in U_r(t)\right\} \supset B(\theta_t(x_*(t),u_*(t)),\delta) \text{ a.e.}$$

The proof of Theorem 3.1 adapts to give:

Theorem 3.2. *If in the context of Theorem 3.1 we replace hypothesis* [H5] *by* [H5]$_{\mathbf{T}}$*, where Φ_t is a closed convex cone for each t, then we obtain the same conclusions.*

The only modification required is to now choose $u_1^+(t) \in U_r(t)$ to satisfy

$$\theta_t(x_*(t),u_1^+(t)) = \delta e_1 + \tau_1(t), \text{ where } \tau_1(t) \in \Phi_t,$$

and so on. This leads as before to (3.4), since (for example) $\lambda \cdot \tau_1 \le 0$, as a result of the fact that $\lambda \in N^C_{\Phi_t}(\phi_t(x,c))$; the resulting nonpositive term is simply absorbed in γ_1 by redefining the latter. The proof then proceeds as before.

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. The maximum principle in optimal control. *Control and Cybernetics*, 34:709–722, 2005.

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

[5] F. H. Clarke and MdR. de Pinho. Optimal control problems with mixed constraints. Preprint, 2009. (Under revision for SICOPT.)

[6] 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.

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

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

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

[10] MdR. 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.

[11] MdR. 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.

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

[13] 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.
[14] 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.
[15] M. R. Hestenes. *Calculus of Variations and Optimal Control Theory.* Wiley, New York, 1966.
[16] 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.
[17] A. A. Milyutin and N. P. Osmolovskii. *Calculus of Variations and Optimal Control.* American Math. Soc., Providence, 1998.
[18] L. W. Neustadt. *Optimization.* Princeton University Press, Princeton, 1976.
[19] Z. Páles and V. Zeidan. First- and second-order necessary conditions for control problems with constraints. *Trans. Amer. Math. Soc.*, 346:421–453, 1994.
[20] 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.
[21] Z. Páles and V. Zeidan. Optimal control problems with set-valued control and state constraints. *SIAM J. Optim.*, 14:334–358, 2003.
[22] A. B. Schwarzkopf. Relaxed control problems with state equality constraints. *SIAM J. Control*, 13:677–694, 1975.
[23] A. B. Schwarzkopf. Optimal controls with equality state constraints. *J. Optim. Th. Appl.*, 19:455–468, 1976.
[24] G. Stefani and P. Zezza. Optimality conditions for a constrained control problem. *SIAM J. Control Optim.*, 34:635–659, 1996.