J. CONTROL AND OPTIMIZATION
Vol. 14, No. 5, August 1976

LAGRANGE DUALITY THEORY FOR CONVEX CONTROL PROBLEMS*

WILLIAM W. HAGER† AND SANJOY K. MITTER‡

Abstract. The Lagrange dual of control problems with linear dynamics, convex cost and convex inequality state and control constraints is analyzed. If an interior point assumption is satisfied, then the existence of a solution to the dual problem is proved; if there exists a solution to the primal problem, then a complementary slackness condition is satisfied. A necessary and sufficient condition for feasible solutions in the primal and dual problems to be optimal is also given. The dual variables p and v corresponding to the system dynamics and state constraints are proved to be of bounded variation while the multiplier corresponding to the control constraints is proved to lie in $\mathscr{L}^1$. Finally, a control and state minimum principle is proved. If the cost function is differentiable and the state constraints have two derivatives, then the state minimum principle implies that a linear combination of p and v satisfy the conventional adjoint condition for state constrained control problems.

1. Introduction. The Lagrange dual of the following control problem is studied:

$$\inf c(x, u)$$

$$\text{subject to } \dot{x}(t) = A(t)x(t) + B(t)u(t), \qquad x(0) = x_0,$$

$$K_c(u(t), t) \leqq 0, \qquad K_s(x(t), t) \leqq 0,$$

where $c(\cdot, \cdot)$, $K_c(\cdot, t)$ and $K_s(\cdot, t)$ are all convex. Rockafellar [7] has derived duality results for convex state constrained control problems using Fenchel duality theory. The development in this paper goes beyond Rockafellar's results since the constraints are given explicitly by inequalities above, and hence the multipliers associated with the constraints can be characterized. Also, a slightly different form of the dual problem, the Lagrange dual, is studied herein; and the matrix $B(t)$ above, which Rockafellar assumes is the identity matrix in his development, is introduced. The theory in this paper provides the foundation for an analysis of the numerical solution of the dual problem by the Ritz method in [1]. The control problem stated above involves no constraints on $x(0)$ and $x(1)$ except for the condition $x(0) = x_0$; however, convex inequality and linear equality endpoint constraints could have been included with very little change in the analysis. To keep the presentation simpler, these constraints are not explicitly treated; however, notice that the state constrained problem explicitly involves endpoint restrictions because of the condition $K_s(x(t), t) \leqq 0$ for all $t \in [0, 1]$.

In §§2 and 3, the principal result based on the Hahn–Banach theorem, proves that the dual problem has an optimal solution if there exists an interior point for the constraint set (i.e., the Slater condition holds); if the primal problem has an optimal solution, then a complementary slackness condition holds. The optimal multipliers $\hat{p}$ and $\hat{v}$ corresponding to the system dynamics and state

* Received by the editors August 22, 1974, and in revised form May 15, 1975.

† Department of Mathematics, University of South Florida, Tampa. Florida 33620.

‡ Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

The second author was supported in part by the National Science Foundation under Grant GK 41647, by the U.S. Air Force under Grant AFOSR 72 2273, and by NASA under Grant NASA/AMES NGL 22-009-124.

constraints are shown to have bounded variation while the multiplier $\hat{w}$ corresponding to the control constraints lies in L^1. Also a necessary and sufficient condition for the optimality of solutions to the primal and the dual problem is given.

Section 4 then proves that a minimum principle holds, and while $(\hat{p}, \hat{v})$ are only of bounded variation, the combination $\hat{q}(t) = K_s(\hat{x}(t), t)_x^T \hat{v}(t) - \hat{p}(t)$ is absolutely continuous where $\hat{x}$ solves the primal problem; furthermore $\hat{q}$ satisfies the conventional adjoint equation for state constrained control problems. This result has important consequences for the solution of the dual problem using the Ritz method in [1] since the convergence rate of the discrete approximation depends upon the smoothness of the dual variables; hence if the dual problem is reformulated in terms of q rather than p, then a superior convergence rate is achieved.

The Appendix contains several technical lemmas concerning the regularity of the dual variables.

Notation. The following notation is used for spaces of real-valued functions on [0, 1]:

$\mathcal{A}$	absolutely continuous functions,
$\mathcal{BV}$	functions of bounded variation continuous from the left on [0, 1),
$\mathcal{NBV}$	functions of bounded variation continuous from the left on [0, 1), and normalized so that $f(1) = 0$,
$\mathcal{C}$	continuous functions,
$\mathcal{L}^p$	functions with $\int_0^1 \|f(t)\|^p \, dt < \infty$,
$\mathcal{L}^\infty$	functions essentially bounded and measurable.

If $\mathcal{W}$ is any of the spaces above, the notation $x \in \mathcal{W}(R^n)$ means that x is a vector-valued function with n components and each component lies in $\mathcal{W}$.

If $y \in R^m$, then define $|y| = \sum_{k=1}^m |y_k|$ and denote the supremum norm of a vector-valued function by $\|f\| = \sup_{t \in [0,1]} |f(t)|$.

If $x, y \in R^m$, the inner product $(\cdot , \cdot)$ is defined by $(x, y) = \sum_{k=1}^m x_k y_k$. If $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$, where $\mathcal{L}^q$ is the dual of $\mathcal{L}^p$, $v \in \mathcal{BV}$, and $h \in \mathcal{C}$, then define:

$$\langle f, g \rangle = \int_0^1 (f(t), g(t)) \, dt, \qquad [v, h] = \int_0^1 h(t) \, dv(t).$$

The *complement* and *closure* of a set are denoted A^c and $\bar{A}$, respectively.

2. Duality theory.

The following *control problem* is considered:

$$\begin{aligned}
&\inf c(x, u) \\
&\text{subject to } c(x, u) = \int_0^1 h(x(t), u(t), t) \, dt, \\
\text{(P)} \qquad &\dot{x}(t) = A(t)x(t) + B(t)u(t), \qquad x(0) = x_0, \\
&K_c(u(t), t) \leqq 0, \qquad K_s(x(t), t) \leqq 0 \quad \text{for all } t \in [0, 1], \\
&x \in \mathcal{A}(R^n), \qquad u \in \mathcal{L}^\infty(R^m),
\end{aligned}$$

where h, K_c and K_s have range in R, R^{m_c} and R^{m_s}, respectively, and the matrices A and B are of the appropriate dimensions. Note that in the control problem above, the controls lie in $\mathscr{L}^\infty$. The next section will treat the case where the controls lie in $\mathscr{L}^1$. The *dual function L* is given by

$$L(p, w, v) = \inf \{c(x, u) + \langle p, \dot{x} - Ax - Bu\rangle + \langle w, K_c(u)\rangle + [v, K_s(x)]\}$$
(1) $$\text{subject to } x(0) = x_0, \quad x \in \mathscr{A}(R^n), \quad u \in \mathscr{L}^\infty(R^m).$$

The dual problem corresponding to (P)

$$\sup L(p, w, v)$$
(D) $$\text{subject to } p \in \mathscr{BV}(R^n), \quad v \in \mathscr{NBV}(R^{m_s}), \quad w \in \mathscr{L}^1(R^{m_c}), \quad w \geqq 0,$$
$$v \text{ nondecreasing.}$$

In order that all the terms in (P) and (1) above make sense, assumptions must be made concerning the functions appearing in these problems. Theorem 1 will require the following, continuity, convexity and Slater conditions:

(C) $h(\cdot, \cdot, t)$, $K_s(\cdot, t)$ and $K_c(\cdot, t)$ are convex for $t \in [0, 1]$, $A(\cdot)$ and $B(\cdot)$ have components in $\mathscr{L}^1$ and $h(\cdot, \cdot, \cdot)$, $K_s(\cdot, \cdot)$ and $K_c(\cdot, \cdot)$ are all continuous.

(SL) There exists a control $\bar{u} \in \mathscr{C}(R^m)$ and a corresponding trajectory $\bar{x}$ such that $(K_c(\bar{u}(t), t))_j < a < 0$ and $(K_s(\bar{x}(t), t))_j < a < 0$ for some "a", for all $t \in [0, 1]$ and for all components of K_c and K_s.

Proposition 1 below, the *weak duality theorem*, is easily verified. This is followed by the principal theorem, or *strong duality* result.

PROPOSITION 1. *$c(x, u) \leqq L(p, w, v)$ whenever (x, u) are feasible in* (P) *and (p, w, v) are feasible in* (D).

THEOREM 1. *Suppose* (C) *and* (SL) *hold and the optimal value, $\hat{c}$, of* (P) *is finite. Then there exist $(\hat{p}, \hat{w}, \hat{v})$ that are optimal in* (D) *with $L(\hat{p}, \hat{w}, \hat{v}) = \hat{c}$. Furthermore, if $(\tilde{p}, \tilde{w}, \tilde{v})$ and $(\tilde{x}, \tilde{u})$ are feasible in* (D) *and* (P), *respectively, then a necessary and sufficient condition for $(\tilde{p}, \tilde{w}, \tilde{v})$ and $(\tilde{x}, \tilde{u})$ to be optimal solutions to the dual and primal problems is that $(\tilde{x}, \tilde{u})$ achieve the minimum in* (1) *for $(p, w, v) = (\tilde{p}, \tilde{w}, \tilde{v})$ and the complementary slackness conditions $\langle \tilde{w}, K_c(\tilde{u})\rangle = [\tilde{v}, K_s(\tilde{x})] = 0$ hold.*

Observe that the condition $\langle \tilde{w}, K_c(\tilde{u})\rangle = [\tilde{v}, K_s(\tilde{x})] = 0$ implies that $K_c(\tilde{u}(t), t)_j = 0$ whenever $\tilde{w}(t)_j > 0$ a.e. and $\tilde{v}_j$ is constant on every interval where $K_s(\tilde{x}(t), t) < 0$. Also notice that the sufficiency condition follows immediately from complementary slackness, feasibility of $(\tilde{x}, \tilde{u})$ and $(\tilde{p}, \tilde{w}, \tilde{v})$, the optimality of $(\tilde{x}, \tilde{u})$ in (1) for $(p, w, v) = (\tilde{p}, \tilde{w}, \tilde{v})$ and Proposition 1; that is, $c(\tilde{x}, \tilde{u}) = L(\tilde{p}, \tilde{w}, \tilde{v})$, and this can only happen when $(\tilde{x}, \tilde{u})$ and $(\tilde{p}, \tilde{w}, \tilde{v})$ are optimal in (P) and (D), respectively. On the other hand, if $(\tilde{p}, \tilde{w}, \tilde{v})$ and $(\tilde{x}, \tilde{u})$ are optimal in (D) and (P) and it can be proved that the optimal value of the primal and dual problems are equal, then $c(\tilde{x}, \tilde{u}) = L(\tilde{p}, \tilde{w}, \tilde{v}) \leqq c(\tilde{x}, \tilde{u}) + \langle \tilde{w}, K_c(\tilde{u})\rangle + [\tilde{v}, K_s(\tilde{x})]$. Since $K_c(\tilde{u}(t), t) \leqq 0$, $\tilde{w}(t) \geqq 0$, $K_s(\tilde{x}(t), t) \leqq 0$ and $\tilde{v}$ is nondecreasing, $\langle \tilde{w}, K_c(\tilde{u})\rangle = 0$, $[\tilde{v}, K_s(\tilde{x})] = 0$ and $(\tilde{x}, \tilde{u})$ achieve the minimum in (1) for $(p, w, v) = (\tilde{p}, \tilde{w}, \tilde{v})$. Thus the proof of Theorem 1 will be complete if it can be shown that the optimal value of the dual problem and the primal problem are equal whenever (SL) and (C) hold and the value of the primal problem is finite.

Rather than prove directly that the optimal value of the primal and dual problem are equal, we first consider a slightly more general problem:

$$\begin{aligned}&\inf f(x,u)\\ (\mathrm{P}')\quad &\text{subject to } \dot{x}(t)=A(t)x(t)+B(t)u(t),\quad x(0)\in X_0,\quad K_s(x(t),t)\leqq 0,\\ &u(t)\in U(t)\quad \text{for all } t\in[0,1],\quad x\in\mathscr{A}(R^n),\quad u\in\mathscr{L}^\infty(R^m),\end{aligned}$$

where f is a functional defined on $\mathscr{A}(R^n)\times\mathscr{L}^\infty(R^m)$. The corresponding dual function is

$$\begin{aligned}&L'(p,v)=\inf\{f(x,u)+\langle p,\dot{x}-Ax-Bu\rangle+[v,K_s(x)]\}\\ (2)\quad &\qquad \text{subject to } x\in\mathscr{A}(R^n),\quad u\in\mathscr{L}^\infty(R^m),\quad x(0)\in X_0,\quad u(t)\in U(t)\\ &\qquad\qquad \text{for all } t\in[0,1].\end{aligned}$$

The dual problem is

$$\begin{aligned}&\sup L'(p,v)\\ (\mathrm{D}')\quad &\text{subject to } p\in\mathscr{BV}(R^n),\quad v\in\mathscr{NBV}(R^{m_s}),\quad v \text{ nondecreasing}.\end{aligned}$$

Define $X=\{x\in\mathscr{A}(R^n): K_s(x(t),t)\leqq 0$ for all $t\in[0,1]\}$ and $U=\{u\in\mathscr{L}^\infty(R^m);\ u(t)\in U(t)$ for all $t\in[0,1]\}$, and make the following assumptions analogous to those above for problem (P).

(C′) $f(\cdot,\cdot)$, $K_s(\cdot,t)$, $U(t)$ and X_0 are convex for all $t\in[0,1]$, $K_s(\cdot,\cdot)$ is continuous, and both $A(\cdot)$ and $B(\cdot)$ have components in $\mathscr{L}^1$.

(SL′) There exists a control $\bar{u}\in\mathscr{C}(R^m)$, a corresponding trajectory $\bar{x}$ and constants M, ρ, $\alpha>0$ such that $\bar{u}\in U$, $\bar{x}(0)\in X_0$, $K_s(\bar{x}(t),t)_j<-\alpha<0$ for all components of K_s, and $f(x,\bar{u})<M$ whenever $\|x-\bar{x}\|\leqq\rho$.

LEMMA 1. *Suppose* (C′) *and* (SL′) *hold and* $\hat{c}$, *the optimal value of* (P′), *is finite. Then there exist* $(\hat{p},\hat{v})$ *that are optimal in* (D′) *and* $L'(\hat{p},\hat{v})=\hat{c}$. *If* $(\hat{x},\hat{u})$ *are optimal in* (P′), *then* $[\hat{v},K_s(\hat{x})]=0$ *and hence* $(\hat{x},\hat{u})$ *achieve the minimum in* (2) *for* $(\hat{p},\hat{v})$.

Proof. Lemma 1 follows from an application of the Hahn–Banach theorem to the following two sets:

$$\begin{aligned}Y=\{(a,b,c): &\ a\in R^1, b\in\mathscr{L}^1(R^n), c\in\mathscr{C}(R^{m_s}), a\leqq\hat{c}, b=0, c\leqq 0\},\\ Z=\{(a,b,c): &\ a\in R^1, b\in\mathscr{L}^1(R^n), c\in\mathscr{C}(R^{m_s}) \text{ and there exists}\\ &x\in\mathscr{A}(R^n) \text{ and } u\in U \text{ with } x(0)\in X_0, a\geqq f(x,u),\\ &b(t)=\dot{x}(t)-A(t)x(t)-B(t)u(t), c(t)\geqq K_s(x(t),t) \text{ for all}\\ &t\in[0,1]\}.\end{aligned}$$

From the development of duality in the literature, it is obvious that two sets like Y and Z must be constructed, and the hyperplane separating the sets will define the optimal dual multipliers. Note though that the choice of the convex sets that are to be separated is a very delicate question since one set must have nonempty *interior* which is *disjoint* from the other set before the Hahn–Banach theorem can be employed. Also the sets must be chosen so that the dual multipliers are in "reasonable" spaces if the duality principle corresponding to the sets is to generate a *numerically tractable problem.* It will be seen that Y and Z do indeed satisfy all these conditions and lead to the duality principle stated in the lemma.

The reader can readily verify that the convexity conditions in (C′) imply that Y and Z are convex, the assumption (SL′) implies that Z has an interior point, and the fact that $\hat{c}$ is the optimal value in (P′) implies that Y and the interior of Z are disjoint. Thus by the Hahn–Banach theorem [4], there exists a hyperplane separating Z and Y, i.e., there exists $r \in R^1$, $p \in \mathscr{L}^\infty(R^n)$, $v \in \mathcal{NBV}(R^{m_s})$ such that

$$(r, a_1)+\langle p, b_1\rangle+[v, c_1] \geqq (r, a_2)+\langle p, b_2\rangle+[v, c_2] \tag{3}$$

for all $(a_1, b_1, c_1) \in \bar{Z}$, $(a_2, b_2, c_2) \in \bar{Y}$. By choosing particular points in Y and Z, properties of the separating hyperplane will be exhibited:

(a) $r \geqq 0$. Substitute $a_2 = \hat{c} - 1$, $a_1 = f(\bar{x}, \bar{u})$, $b_1 = b_2 = c_1 = c_2 = 0$ in (3) where $(\bar{x}, \bar{u})$ was given in (SL′).

(b) *v is monotone nondecreasing.* For notational convenience, v is assumed scalar-valued, although for vector-valued functions the proof is identical.

Given $t, s, d \in [0, 1)$, $t < s$, $d < |s-t|$, let c_d denote the continuous piecewise linear function that is -1 on $[t, s-d]$, zero on $[0, t-d]$ and $[s, 1]$, and linear on $[t-d, t]$ and $[s-d, s]$. Now, $[v, c_d] = v(t) - v(s-d) + z_d$ where

$$|z_d| \leqq |TV(t, v) - TV(t-d, v)| + |TV(s, v) - TV(s-d, v)|$$

and $TV(t, v)$ is the total variation of v on $[0, t]$. Since v is continuous from the left on $[0, 1)$, then $TV(\cdot, v)$ is continuous from the left at t and s (see [6]), and hence $\lim_{d\to 0} |z_d| = 0$ and $\lim_{d\to 0} [v, c_d] = v(t) - v(s)$. Substituting $(\hat{c}, 0, 0) \in \bar{Z}$ and $(\hat{c}, 0, c_d) \in Y$ into (3) and letting $d \to 0$, we obtain $v(t) \leqq v(s)$. The right endpoint, $t = 1$, is treated similarly.

(c) *If* $(\hat{x}, \hat{u})$ *are optimal in* (P′), *then* $[v, K_s(\hat{x})] = 0$. Substitute $a_1 = a_2 = \hat{c}$, $b_1 = b_2 = c_2 = 0$, and $c_1(t) = K_s(\hat{x}(t), t)$ in (3). Then $[v, K_s(\hat{x})] \geqq 0$ and (c) follows from (b). Hence the complementary slackness condition in the lemma holds.

(d) $r > 0$. Suppose $r = 0$. Substituting $b_1 = b_2 = c_2 = 0$ and $c_1(t) = K_s(\bar{x}(t), t)$ in (3) yields $[v, K_s(\bar{x})] \geqq 0$. Since $K_s(\bar{x}(t), t)_j < -a < 0$, (b) implies that $v = 0$. Substituting $b_1 = -p$ and $b_2 = 0$ in (3) yields $-\langle p, p\rangle \geqq 0$. Hence, $p = 0$ a.e. This is impossible since r, p, v cannot all vanish so that $r > 0$ and (3) can be normalized with $r = 1$.

(e) $L'(p, v) = \hat{c}$. Substituting $a_1 = c(x, u)$, $b_1 = \dot{x} - Ax - Bu$, $c_1 = K_s(x)$, $a_2 = \hat{c}$, $b_2 = c_2 = 0$ in (3) and recalling that $r = 1$ from (d) yields $L'(p, v) \geqq \hat{c}$. However, by weak duality, $L'(p, v) \leqq \hat{c}$ and hence $L'(p, v) = \hat{c}$. Note that $p \in \mathscr{L}^\infty(R^n)$, but the lemma claims that $L'(p, v) = \hat{c}$ where $p \in \mathcal{BV}$.

(f) $p = \tilde{p}$ a.e. where $\tilde{p} \in \mathcal{BV}$. This proof is more technical than (a) to (e) and appears in Lemma A. 1 of the Appendix, so the proof of Lemma 1 is complete since $L'(p, v) = L'(\tilde{p}, v)$. □

Proof of Theorem 1. In the problem (P) with explicit control constraints, proceed exactly as in the proof of Lemma 1. A fourth component $d \in \mathscr{C}(R^{m_c})$ is added to the sets Y and Z, where $d \leqq 0$ in Y and $d(t) \geqq K_c(u(t), t)$ in Z. (Note that $d \in \mathscr{C}$ and not $d \in \mathscr{L}^\infty$—if d were chosen in $\mathscr{L}^\infty$, then the Hahn–Banach theorem would produce a multiplier in the dual of $\mathscr{L}^\infty$ which is a miserable space. By choosing $d \in \mathscr{C}$, the dual multiplier lies in $\mathscr{NBV}$— in fact, it is seen below that the multiplier is also absolutely continuous.)

Continuing as in Lemma 1, we find the Hahn–Banach theorem yields

$$c(x, u)+\langle p, \dot{x}-Ax-Bu\rangle+[v, K_s(x)]+[z, K_c(u)] \geqq \hat{c} \tag{4}$$

for all $x \in \mathscr{A}(R^n)$ with $x(0)=x_0$ and $u \in \mathscr{C}(R^m)$ where $v \in \mathscr{NBV}(R^{m_s})$, $z \in \mathscr{NBV}(R^{m_c})$, and both v and z are nondecreasing. Note that to obtain an optimal solution to (D), it must be shown that: (i) z is absolutely continuous so that $[z, K_c(u)]=\langle w, K_c(u)\rangle$ where $w=\dot{z}$ and (ii) expression (4) holds for all $u \in \mathscr{L}^\infty(R^m)$, not just for $u \in \mathscr{C}(R^m)$. Combining these properties with weak duality, Proposition 1, implies that $L(p, w, v)=\hat{c}$.

First it is proved that the infimum of the left side of (4) over $x \in \mathscr{A}(R^n)$ and $u \in \mathscr{C}(R^m)$ actually equals $\hat{c}$. Let $\{u^k\}$ be a minimizing sequence for (P) and let $\{x^k\}$ be the corresponding trajectories. The sequence $\{u^k\}$ lies in $\mathscr{L}^\infty$; however, in Lemma A.2 of the Appendix, it is shown that the convexity of K_c and the existence of an interior point for the constraint $K_c(u(t), t) \leqq 0$ (given in (SL)) imply that for any $\varepsilon>0$, there exists $y_\varepsilon^k \in \mathscr{C}(R^m)$ satisfying $K_c(y_\varepsilon^k(t), t) \leqq 0$, $|y_\varepsilon^k(t)-u^k(t)| \leqq \varepsilon$ except on a set of measure less than ε, and $\|y_\varepsilon^k\| \leqq \|\bar{u}\|+\|u^k\|$, where $\bar{u}$ is the interior control given in (SL). Thus, by the continuity of $h(\cdot,\cdot,\cdot)$, the integrand of the cost functional of (P), it follows that $\lim_{\varepsilon\to 0} c(x^k, y_\varepsilon^k)=c(x^k, u^k)$ and $\lim_{\varepsilon\to 0}\langle p, \dot{x}^k-Ax^k-By_\varepsilon^k\rangle=0$. Now given $\delta>0$, there exist k' such that $|c(x^{k'}, u^{k'})-\hat{c}|<\delta/3$ and ε' such that $|c(x^{k'}, y_{\varepsilon'}^{k'})-c(x^{k'}, u^{k'})|<\delta/3$ and $|\langle p, \dot{x}^{k'}-Ax^{k'}-By_{\varepsilon'}^{k'}\rangle|<\delta/3$. Since $[z, K_c(y_{\varepsilon'}^{k'})] \leqq 0$ and $[v, K_s(x^{k'})] \leqq 0$, then the left side of (4) evaluated at $x=x^{k'}$ and $u=y_{\varepsilon'}^{k'}$ is within δ of $\hat{c}$, and hence the infimum of the left side over (x, u) satisfying $x \in \mathscr{A}(R^n)$, $u \in \mathscr{C}(R^m)$ and $x(0)=x_0$ equals $\hat{c}$ as claimed.

The proof that $z \in \mathscr{A}(R^{m_c})$ is now summarized, and the details can be found in Lemma A.3 of the Appendix.

Define $g(x, u)=c(x, u)+\langle p, \dot{x}-Ax-Bu\rangle+[v, K_s(x)]$. Using the construction of the previous paragraph, there exists a sequence (x^k, y^k) satisfying $g(x^k, y^k)\to\hat{c}$, $[z, K_c(y^k)] \leqq 0$, and $y^k \in \mathscr{C}(R^m)$. It is possible to express $z=r+s$ where $r \in \mathscr{A}(R^{m_c})$, $s \in \mathscr{BV}(R^{m_c})$, $\dot{s}=0$ a.e., $s(0)=0$ and s is nondecreasing (see Rudin [8, p. 166]). In Lemma A.3 of the Appendix, it is shown that a sequence $\{\delta^k\} \subset \mathscr{C}(R^{m_c})$ can be constructed with $\delta^k=0$ except on a set E of small measure on which is concentrated the variation on s, $\delta^k=\bar{u}-y^k$ just inside E, and hence $[s, K_c(y^k+\delta^k)] \leqq as(1)/2$ where $a<0$ was given in (SL). Since s is nondecreasing, then $s(1) \geqq 0$, and unless $s=0$, (4) will be contradicted since $g(x^k, y^k+\delta^k)+as(1)/2$ will be less than $\hat{c}$ for k sufficiently large. Hence $z=r \in \mathscr{A}(R^{m_c})$.

To complete the proof, it must be shown that (4) holds for $u \in \mathscr{L}^\infty(R^m)$, not just $u \in \mathscr{C}(R^m)$. By Lusin's theorem [8, p. 53], any $u \in \mathscr{L}^\infty(R^m)$ can be approximated by $y_\varepsilon \in \mathscr{C}(R^m)$ satisfying $y_\varepsilon=u$ except on a set of measure less than ε and $\|y_\varepsilon\| \leqq \|u\|$. Since (4) holds for y_ε, the continuity condition (C) implies that (4) holds for $u \in \mathscr{L}^\infty(R^m)$. Thus $L(p, w, v)=\hat{c}$ as desired and the complementary slackness conditions follow as in Lemma 1, property (c). □

Notice that the duality results above were derived by separating the sets Y and Z with a hyperplane, and exploiting the separation condition (4) above to push the dual variables into successively smaller spaces. An immediate question is whether the spaces exhibited above are the smallest possible. A more recent paper [11] will show that for a strictly convex, quadratic cost control problem with linear state and control constraints satisfying an independence condition, there exists an optimal control u^*, a corresponding trajectory x^* and dual multipliers p^*, w^* and v^* such that $(\dot{x}^*, u^*, p^*, w^*, v^*)$ are all Lipschitz continuous when K_s has a Lipschitz continuous partial derivative in t and A, B, K_c, and h are Lipschitz continuous in t. Furthermore, if no state constraints are present, then $\ddot{p}^*$ is Lipschitz continuous when the data defining (P) is sufficiently smooth. Below it is shown that when state constraints are present, a linear combination, q^*, of p^* and v^* has increased smoothness, and in [11] the Lipschitz continuity of $\dot{q}^*$ is proved. Hence $(\dot{x}^*, \dot{q}^*, u^*, w^*, v^*)$ have derivatives in L^∞. Also by an example given in [11], it is seen that $(\dot{x}^*, \dot{q}^*, u^*, w^*, v^*)$ may be discontinuous when K_s does not possess a Lipschitz continuous partial derivative t.

3. Extension of duality theory to controls in $\mathscr{L}^1$. Let $(\bar{\mathrm{P}})$ denote the control problem with constraint $u \in \mathscr{L}^1(R^m)$ instead of $u \in \mathscr{L}^\infty(R^m)$. It is assumed both that the components of $B(\cdot)$ lie in $\mathscr{L}^\infty$ so that the differential equation $\dot{x} = Ax + Bu$ makes sense, and the integral in the cost functional is defined for $x \in \mathscr{A}(R^n)$ and $u \in \mathscr{L}^1(R^m)$ (i.e., the integrand is in $\mathscr{L}^1$).

THEOREM 2. *Suppose* (C) *and* (SL) *hold and the optimal value* $\bar{c}$ *of* $(\bar{\mathrm{P}})$ *is finite. Then there exist* $(\hat{p}, \hat{w}, \hat{v})$ *that are optimal in* (D) *with* $L(\hat{p}, \hat{w}, \hat{v}) = \bar{c}$. *If* $(\hat{x}, \hat{u})$ *are optimal in* $(\bar{\mathrm{P}})$, *then the complementary slackness condition of Theorem* 1 *holds.*

Note that in defining the dual problem (D), we still restrict $u \in \mathscr{L}^\infty(R^m)$ in the minimization of (1).

Proof. Let $\hat{c}$ denote the optimal value of (P). Since $\hat{c} \geqq \bar{c} > -\infty$, then Theorem 1 implies the existence of (p, w, v) with

$$c(x, u) + \langle p, \dot{x} - Ax - Bu\rangle + \langle w, K_c(u)\rangle + [v, K_s(x)] \geqq \hat{c} \geqq \bar{c} \tag{5}$$

for all (x, u) satisfying $x \in \mathscr{A}(R^n)$, $x(0) = x_0$ and $u \in \mathscr{L}^\infty(R^m)$. It is now shown that $\hat{c} = \bar{c}$. Suppose for the moment that there exists an optimal solution $(\hat{x}, \hat{u})$ to $(\bar{\mathrm{P}})$. Define the following control u_k and set S_k:

$$u_k = \begin{cases} \hat{u}(t) & \text{when } |\hat{u}(t)| \leqq k, \\ \bar{u}(t) & \text{when } |\hat{u}(t)| > k, \end{cases} \qquad S_k = \{t : \hat{u}(t) \neq u_k(t)\},$$

where $\bar{u}$ was given in (SL).

Since $w \geqq 0$, v is nondecreasing, and $K_c(u_k(t), t) \leqq 0$ and $K_s(\hat{x}(t), t) \leqq 0$ for $t \in [0, 1]$, then inserting $(x, u) = (\hat{x}, u_k)$ into (5) yields $c(\hat{x}, u_k) + \langle p, B(\hat{u} - u_k)\rangle \geqq \hat{c} \geqq \bar{c}$. Since the components of $B(\cdot)$ and $p(\cdot)$ lie in $\mathscr{L}^\infty$, $u \in \mathscr{L}^1(R^m)$, $\hat{u} = u_k$ except on S_k, and $\mu(S_k) \to 0$ as $k \to \infty$, where $\mu(\cdot)$ denotes Lebesque measure, then $0 = \lim_{k\to\infty} \langle p, B(\hat{u} - u_k)\rangle$. Similarly $c(\hat{x}, \hat{u}) - c(\hat{x}, u_k) = \int_{S_k} \{h(\hat{x}(t), \hat{u}(t), t) - h(\hat{x}(t), \bar{u}(t), t)\}\, dt$ and both $h(\hat{x}(\cdot), \hat{u}(\cdot), \cdot)$ and $h(\hat{x}(\cdot), \bar{u}(\cdot), \cdot)$ lie in $\mathscr{L}^1$, so $c(\hat{x}, \hat{u}) = \lim_{k\to\infty} c(\hat{x}, u_k)$. Thus $\hat{c} = \bar{c}$ since the left side of (5) evaluated at $x = \hat{x}$ and $u = u_k$ converges to $\bar{c}$. Since $L(p, w, v) = \hat{c}$, $L(p, w, v) = \bar{c}$.

If there does not exist an optimal solution to (P), then by choosing a minimizing sequence and approximating each element of the minimizing sequence as above, it can be proved that $L(p, w, v)=\hat{c}=\bar{c}$.

Now the complementary slackness condition is verified. Again by the inequality (5) above, substituting $u=u_k$, $x=\hat{x}$ yields:

$$\int_{S_k^c}(w(t), K_c(\hat{u}(t), t))\,dt \geqq \langle w, K_c(u_k)\rangle \geqq \bar{c}-c(\hat{x}, u_k)-\langle p, B(\hat{u}-u_k)\rangle. \tag{6}$$

As shown above, the right side of (6) converges to zero as $k\to\infty$. Since $\lim_{k\to\infty}\mu(S_k^c)=1$, $E_j=\{t: w(t)_j>0,\ K_c(\hat{u}(t), t)_j<0\}$ has no measure, and the complementary slackness condition in the control constraint must hold. A similar proof confirms the complementary slackness condition in the state constraint. □

4. Minimum principles. In order to solve the dual problem numerically, the x and u that achieve the infimum in (1) must be characterized. This leads to a *minimum principle* and an *adjoint condition.* Theorem 3 below proves that the minimization over u in (1) can be taken under the integral sign.

THEOREM 3. *Suppose* (C) *and* (SL) *hold,* (p, w, v) *is feasible in* (D) *with* $L(p, w, v)>-\infty$, *and* $x^*\in\mathcal{A}(R^n)$ *and* $u^*\in\mathcal{L}^\infty(R^m)$ *achieve the minimum in* (1) *corresponding to* (p, w, v). *Then the minimum of* $f(u, t)= h(x^*(t), u, t)-(p(t), B(t)u)+(w(t), K_c(u, t))$ *occurs at* $u=u^*(t)$ *for almost every* $t\in[0, 1]$. *Similarly, if* $L'(p, v)>-\infty$, *the cost functional in* (P′) *is given by* $c(\cdot,\cdot)$, $U(t)=\{b\in R^m: K_c(b, t)\leqq 0\}$, *and* $x^*\in\mathcal{A}(R^n)$ *and* $u^*\in\mathcal{L}^\infty(R^m)$ *achieve the minimum in* (2) *corresponding to* (p, v), *then the minimum of* $\{h(x^*(t), u, t)-(p(t), B(t)u)\}$ *over* $u\in U(t)$ *occurs at* $u=u^*(t)$ *for almost every* $t\in[0, 1]$.

Proof. Only the first minimum principle above will be proved since the second is similar. Let $\bar{c}=L(p, w, v)$ where by definition

$$\begin{aligned} L(p, w, v)=\inf\Big[\int_0^1 &\{h(x(t), u(t), t)+(p(t), \dot{x}(t)-A(t)x(t)-B(t)u(t))\\ &+(w(t), K_c(u(t), t))\}\,dt+[v, K_s(x)]\Big] \\ &\text{subject to } x\in\mathcal{A}(R^n),\quad u\in\mathcal{L}^\infty(R^m),\quad x(0)=x_0. \end{aligned}\tag{7}$$

Let E denote the intersection of the Lebesgue points of each term in the integrand of (7) evaluated at (x^*, u^*) and suppose $f(z, s)<f(u^*(s), s)$ for some $s\in E$ and $z\in R^m$. Let Δ denote a ball of diameter δ centered at s, $I(\Delta, u)$ the integral in (7) evaluated at $x=x^*$ over the ball Δ, and $J(u(\cdot))$ the integrand in (7) evaluated at $x=x^*$. Since s is a Lebesgue point of $J(u^*(\cdot))$, $I(\Delta, u^*)=J(u^*(s))\delta+o(\delta)$. Define v_δ to be a control that agrees with u^* outside Δ and equals z inside Δ. It is easy to see that $I(\Delta, v_\delta)=J(z)\delta+o(\delta)$, and since $f(z, s)<f(u^*(s), s)$, $J(z)< J(u^*(s))$ and $I(\Delta, v_\delta)<I(\Delta, u^*)$ for δ sufficiently small. This violates the optimality of (x^*, u^*) in (7) so that the minimum principle holds on E. Since E has full measure, the proof is complete. □

Note that Theorem 3 holds for all (p, w, v) that are feasible in the dual problem, while the standard necessary conditions only hold for some (p, w, v). Also observe that it is not possible to carry out the minimization over x under the integral sign in (1) due to the presence of the $\dot{x}$ term. The following lemma will be needed before the adjoint conditions can be derived.

LEMMA 2. *Suppose* (C) *and* (SL) *hold,* (p, w, v) *is feasible in* (D) *with* $L(p, w, v) > -\infty$, (x^*, u^*) *achieves the minimum in* (1) *for* (p, w, v), $K_s(\cdot, \cdot)$ *is twice continuously differentiable, and* $G(t)$ *denotes the gradient of* $K_s(\cdot, t)$ *evaluated at* $x^*(t)$. *Then if* q *is defined by* $q(1)=0$, $q(t)=G(t)^T v(t)-p(t)$ *for* $t \in (0, 1)$, *and* $q(0)=q(0^+)$, *then* $q \in \mathscr{A}(R^n)$. *If* K_s *is affine, then the existence of* (x^*, u^*) *is not required.*

Proof. By the definition of L,

$$L(p, w, v) \leqq c(x, u)+\langle p, \dot{x}-Ax-Bu\rangle+\langle w, K_c(u)\rangle+[v, K_s(x)] \tag{8}$$

for all $x \in \mathscr{A}(R^n)$ with $x(0)=x_0$ and $u \in \mathscr{L}^\infty(R^m)$. Each term on the right side of (8) is convex and furthermore the $[v, K_s(x)]$-term is differentiable in x. Recall the following standard necessary condition: Suppose v^* solves the problem: minimize $f(v)+g(v)$ subject to $v \in F$ where f, g and F are all convex and f is differentiable. Then v^* satisfies $g(v^*) \leqq g(v)+(d/dv)f'(v^*)(v-v^*)$ for all $v \in F$ (see [3]). Applying this result to the right side of (8) we get

$$\begin{aligned} &c(x^*, u^*)+\langle p, \dot{x}^*-Ax^*-Bu^*\rangle+\langle w, K_c(u^*)\rangle \\ &\qquad \leqq c(x, u)+\langle p, \dot{x}-Ax-Bu\rangle+[v, G(x-x^*)]+\langle w, K_c(u)\rangle \end{aligned} \tag{9}$$

for all $x \in \mathscr{A}(R^n)$ with $x(0)=x_0$ and $u \in \mathscr{L}^\infty(R^m)$. Observe that equality holds in (9) for $x=x^*$ and $u=u^*$.

Since p is continuous from the left on $[0, 1)$, the integration by parts formula of Dunford and Schwartz [4, p. 154] gives

$$\oint_0^1 (p(t), \dot{x}(t)-\dot{x}^*(t))\,dt=(p(1^-), x(1)-x^*(1))-\int_{0^+}^{1^-}[x(t)-x^*(t)]^T\,dp(t). \tag{10}$$

The boundary term at $t=0$ vanishes since $x(0)=x^*(0)=x_0$. Since K_s has two continuous derivatives, then the gradient of $K_s(\cdot, t)$ is absolutely continuous, and hence $G(\cdot)$ is absolutely continuous. Thus the following relation holds:

$$\int_{0^+}^{1^-} x(t)^T G(t)^T\,dv=\int_{0^+}^{1^-} x(t)^T\,d(G(t)^T v)-\int_{0^+}^{1^-} v(t)^T \dot{G}(t)x(t)\,dt. \tag{11}$$

Since v is normalized with $v(1)=0$ and since $x(0)=x^*(0)=x_0$,

$$\begin{aligned} \int_0^1 (x(t)-x^*(t))^T G(t)^T\,dv &= \int_{0^+}^{1^-} (x(t)-x^*(t))^T G(t)^T\,dv \\ &\quad -(x(1)-x^*(1))^T G(1)^T v(1^-). \end{aligned} \tag{12}$$

Combining (9), (10), (11) and (12) we find

$$\begin{aligned} c(x, u)-\langle p, Ax+Bu\rangle-\langle v, \dot{G}x\rangle+\int_{0^+}^{1^-} x(t)^T\,dq(t)+\langle w, K_c(u)\rangle \\ -(q(1^-), x(1)) \geqq \bar{c} > -\infty \end{aligned} \tag{13}$$

for all (x, u) satisfying $x \in \mathscr{A}(R^n)$, $u \in \mathscr{L}^\infty(R^m)$ and $x(0)=x_0$, where $\bar{c} > -\infty$ is a constant depending on x^*, u^*, p, w and v. Again equality holds in (13) for $x=x^*$ and $u=u^*$. If K_s is affine, then (13) holds without even assuming the existence of (x^*, u^*), and $\bar{c}$ only depends on $L(p, w, v)$.

Now it is shown that $q(1^-)=0$. Define the continuous function $g(\delta, \varepsilon, t)$ as follows: $g(\delta, \varepsilon, \cdot)$ is linear on $[1-\varepsilon, 1]$ and satisfies $g(\delta, \varepsilon, t)=0$ for $t\in[0, 1-\varepsilon]$ and $g(\delta, \varepsilon, 1)=\delta q(1^-)$. Inserting $x(t)=x_0+g(\delta, \varepsilon, t)$ into (13) and letting $\varepsilon\to 0$ and $\delta\to+\infty$, we get a contradiction since the left side of (13) deverges to $-\infty$ due to the presence of the boundary term in (13).

Now consider the absolute continuity of q. It is possible to express $q=r+s$, where $r\in\mathscr{A}(R^n)$, $s\in\mathscr{BV}(R^n)$, $s(0)=0$ and $\dot{s}=0$ a.e. (see Rudin [8, p. 166]). If $E=\{t: \dot{s}(t)\neq 0\}$, then Lemma A.4 in the Appendix proves that unless $s=0$, a sequence $\{x_k\}\subset\mathscr{A}(R^n)$ can be chosen such that x_k agrees with $\bar{x}$ just outside of E and $[s, x_k]\to-\infty$. This will violate (13), and hence $s=0$ and $q=r\in\mathscr{A}(R^n)$. □

THEOREM 4. *Suppose* (C) *and* (SL) *hold,* (p, w, v) *is feasible in* (D) *with* $L(p, w, v)>-\infty$, $x^*\in\mathscr{A}(R^n)$ *and* $u^*\in\mathscr{L}^\infty(R^m)$ *achieve the minimum in* (1) *corresponding to* (p, w, v) *and* $K_s(\cdot,\cdot)$ *is twice continuously differentiable. Then the minimum of*

$$f(x, t)=h(x, u^*(t), t)+(\dot{q}(t)+A^T(t)q(t)-(\dot{G}(t)^T+A(t)^TG(t)^T)v(t), x)$$

occurs at $x=x^*(t)$ *for almost every* $t\in[0, 1]$, *where* G *and* q *were defined in Lemma* 2. *If* $h(\cdot, u, t)$ *is differentiable, then the adjoint equation holds*: $q(1)=0$ *and*

$$\dot{q}(t)=-A(t)^Tq(t)-h(x^*(t), u^*(t), t)_x+(\dot{G}(t)^T+A(t)^TG(t)^T)v(t) \quad a.e.\ t \tag{14}$$

Proof. In Lemma 2 it was observed that $q\in\mathscr{A}(R^n)$ so that $[q, x]=\langle\dot{q}, x\rangle$. From (13),

$$\int_0^1 \{h(x(t), u^*(t), t)-(p(t), A(t)x(t))-(v(t), \dot{G}(t)x(t))+(\dot{q}(t), x(t))\}\,dt\geqq\tilde{c} \tag{15}$$

for all $x\in\mathscr{A}(R^n)$ with $x(0)=x_0$, where $\tilde{c}>-\infty$ is a constant depending only on x^*, u^*, p, w and v. As noted after (13), equality holds in (15) for $x=x^*$. As in Theorem 3, we wish to say that $x^*(t)$ yields the pointwise minimum for the integrand. There is one technical point, though, since in Theorem 3, u was contained in $\mathscr{L}^\infty$, while in (15), x lies in $\mathscr{A}$. However, if $z\in R^n$ yields a better minimum for the integrand of (15) at the Lebesgue point $t=s$, then by [10, p. 9] there exists an infinitely differentiable function ϕ_δ^ε that equals 1 on $[s-\delta, s+\delta]$ and equals 0 on $[s+\delta+\varepsilon, 1]$ and $[0, s-\delta-\varepsilon]$. Thus the function $x_\delta^\varepsilon=z\phi_\delta^\varepsilon+(1-\phi_\delta^\varepsilon)x^*$ is absolutely continuous and equals z near s and x^* away from s. Letting first $\varepsilon\to 0$ and then $\delta\to 0$ again violates the optimality of x^*. The adjoint equation is obtained simply by setting the derivative of $f(\cdot, t)$ to zero at $x=x^*(t)$. □

The condition (14) above is the familiar adjoint equation for state constrained problems given in [5] and [2]. These standard necessary conditions only assert that (14) holds for some (p, w, v) where (x^*, u^*) is optimal in (P), while Theorem 4 holds for all (p, w, v) feasible in (D). Using the minimum principles, Theorems 3 and 4 above, the evaluation of $L(p, w, v)$ is reduced to the solution of a sequence of math programming problems for each $t\in[0, 1]$. In certain cases, such as problems with quadratic cost and linear constraints, the minimum principles permit the explicit determination of the (x, u) achieving the minimum in (1) in terms of (p, w, v). The numerical solution of the dual problem using the Ritz method is analyzed in [1].

A combined state and control minimum principle can be proved, and the proof is similar to Theorems 3 and 4 above.

THEOREM 5. *Suppose* (C) *and* (SL) *hold,* (p, w, v) *is feasible in* (D) *with* $L(p, w, v) > -\infty$, $K_s(\cdot, \cdot)$ *is twice continuously differentiable and* $x^* \in \mathcal{A}(R^n)$ *and* $u^* \in \mathcal{L}^\infty(R^m)$ *achieve the minimum in* (1) *corresponding to* (p, w, v). *Then the minimum of* $f(x, u, t)$ *defined below occurs at* $x = x^*(t)$ *and* $u = u^*(t)$ *for a. e.* t:

$$(16)\quad \begin{aligned} f(x, u, t) = {} & h(x, u, t) + (q(t) - G(t)^T v(t), B(t)u) + (w(t), K_c(u, t)) \\ & + (\dot{q}(t) + A(t)^T q(t) - (\dot{G}(t)^T + A(t)^T G(t)^T) v(t), x) \end{aligned}$$

Appendix. Regularity of the dual variables.

LEMMA A.1. *Suppose* (C′) *and* (SL′) *are satisfied, the optimal value* $\hat{c}$ *of* (P′) *is finite, and* $L'(p, v) = \hat{c}$ *where* $p \in \mathcal{L}^\infty(R^n)$ *and* $v \in \mathcal{BV}(R^{m_s})$. *Then* $p = \tilde{p}$ *a.e. where* $\tilde{p} \in \mathcal{BV}(R^n)$.

Proof. For notational convenience, p is assumed scalar-valued (the proof below could be applied to each component of p separately to demonstrate the result for vector-valued functions). Let R denote the set of Lebesgue points and suppose that p has infinite variation on this set. It is now shown that thi to a contradiction.

Given a constant b, there exists $0 = t_0 < t_1 \cdots < t_N$ such that

$$(A.1)\qquad \sum_{1 \leqq j \leqq N, j \text{ odd}} |p(t_{j-1}) - p(t_j)| > b$$

and either $p(t_{j+1}) < p(t_j) > p(t_{j-1})$ for j even or the reverse inequalities hold. For the construction given below, it is assumed that the former holds. Let $\alpha, \rho, M > 0$ be as given in (SL′), and define the function $x_\varepsilon(t)$ as follows: $x_\varepsilon(\cdot)$ is the continuous, piecewise linear function that is zero for j odd and $-\rho$ for j even on the interval $[t_j + \varepsilon, t_{j+1} - \varepsilon]$, linear on the interval $[t_j - \varepsilon, t_j + \varepsilon]$ for all j, and zero at $t = 0$. Notice that as $\varepsilon \to 0$, $\dot{x}_\varepsilon \to -\rho \sum_{j=0}^N (-1)^j \delta(t - t_j)$, where $\delta(\cdot)$ is the delta function, and since $\{t_j\}$ are Lebesgue points of p and $p(t_{j+1}) < p(t_j) > p(t_{j-1})$ for j even, then $\lim_{\varepsilon \to 0} \langle p, \dot{x}_\varepsilon \rangle = \rho \sum_{1 \leqq j \leqq N, j \text{ odd}} p(t_j) - p(t_{j-1}) < -\rho b$.

From the definition of L',

$$(A.2)\qquad f(\bar{x} + x_\varepsilon, \bar{u}) + \langle p, \dot{x}_\varepsilon + \bar{x} - A(\dot{x}_\varepsilon + \bar{x}) - B\bar{u} \rangle + [v, K_s(\bar{x} + x_\varepsilon)] \geqq \hat{c},$$

where $(\bar{x}, \bar{u})$ was given in (SL′). Also by (SL′), $f(\bar{x} + x_\varepsilon, \bar{u}) < M$, and hence all the terms in (A.2) are bound uniformly in b and ε except for the $\langle p, \dot{x}_\varepsilon \rangle$-term which becomes less than $-\rho b$ for ε sufficiently small. Thus if b were chosen sufficiently large, this would lead to a contradiction in (A.2), and hence the total variation of p on R is finite.

Since R has full measure (see [8, p. 158]), for all $t \in R^c$, there exists a sequence $\{t_j\} \subset R$ such that $t_j \to t^-$. Because p has finite variation on R, $\lim_{j \to \infty} p(t_j)$ exists, and it is possible to define a function $\tilde{p}(t)$ that equals $[p(t)]$ for $t \in R$ and equals $[\lim_{j \to \infty} p(t_j)]$ if $t \notin R$ where $\{t_j\} \subset R$ and $t_j \to t^-$. Thus $\tilde{p}(t) = p(t)$ a.e., and $\tilde{p}$ has the same variation on $[0, 1]$ as p has on R. □

The following theorem essentially proves that if the set $U = \{u(\cdot) \in \mathcal{L}^\infty(R^m) : K(u(t), t) \leq 0\}$ has an interior, then any $u(\cdot) \in U$ can be approximated arbitrarily closely in the $\mathcal{L}^p$-norm by a continuous function in U.

LEMMA A.2. *Suppose* $K: R^m \times [0, 1] \to R^n$ *is continuous*, $K(\cdot, t)$ *is convex for* $t \in [0, 1]$, *and there exist* $\bar{u} \in \mathscr{C}(R^m)$ *and* $a < 0$ *such that* $K(\bar{u}(t), t)_j < a$ *for all* $t \in [0, 1]$ *and* $j = 1, 2, \cdots, n$. *Then given* $u(\cdot) \in U$ *and* $\varepsilon > 0$, *there exists* $v \in U \cap \mathscr{C}(R^m)$ *such that* $|u(t) - v(t)| < \varepsilon$ *except on a set of measure less than* ε *and* $\|v\| \leqq \|\bar{u}\| + \|u\|$.

Proof. Let $w = b\bar{u} + (1-b)u$, where $1 > b > 0$ is small enough that $\|u - w\| \leqq \varepsilon$. By the convexity of $K(\cdot, t)$, $K(w(t), t)_j \leqq ba < 0$ for $j = 1, \cdots, n$, and by Lusin's theorem [8, p. 53], there exists $y \in \mathscr{C}(R^m)$ with $y = w$ on a closed set E satisfying $\mu(E^c) \leqq \varepsilon$, where $\mu(\cdot)$ denotes Lebesgue measure and furthermore, $\|y\| \leqq \|w\|$. Since $K(y(\cdot), \cdot)$ is uniformly continuous on $[0, 1]$, there exists a constant $\delta > 0$ such that if $|t - s| < \delta$, then $|K(y(t), t) - K(y(s), s)| < b|a|$. Outer regularity of the Lebesgue measure implies the existence of an open set D containing E with $\mu(D - E) < \delta$. Also D can be chosen so that no point of D is more than δ away from a point of E (for example, construct open balls of diameter δ about each point of E, choose a finite subcover $\{B_j\}$ of the balls, construct an open set $B \supset E$ with $\mu(B - E) < \delta$, and define $D = (\bigcup B_j) \cap B$).

Since $K(y(t), t)_j = K(w(t), t)_j \leqq ba < 0$ on E and any point of D is at most δ away from a point of E, $K(y(t), t) \leqq 0$ for $t \in D$. From Urysohn's lemma, there exists $g \in \mathscr{C}(R^1)$ with the support of g contained in D, $g(t) = 1$ for $t \in E$, and $\|g\| = 1$ (we use the notation of Rudin [8] to denote a function satisfying these conditions: $E < g < D$). Define $v = gy + (1-g)\bar{u}$. For $t \in D$, v is on the line segment between two functions that satisfy the constraint $K(\cdot, t) \leqq 0$, and since $K(\cdot, t)$ is convex, $K(v(t), t) \leqq 0$. On the other hand, $v = \bar{u}$ on D^c so $v \in U$. By construction, $v(t) = y(t) = w(t)$ for $t \in E$ so $|u(t) - v(t)| \leqq \varepsilon$ except on a set E^c of measure less than ε. Also $|v(t)| \leqq g(t)|y(t)| + (1 - g(t))|\bar{u}(t)| \leqq g(t)|w(t)| + (1 - g(t))|\bar{u}(t)| \leqq [g(t)(b-1)+1]|\bar{u}(t)| + (1-b)g(t)|u(t)|$, and since $0 < b < 1$ and $\|g\| = 1$, the bound $\|v\| \leqq \|\bar{u}\| + \|u\|$ is immediate. □

LEMMA A.3. *Suppose* (C) *and* (SL) *hold*; *then the function* $z \in \mathscr{BV}$ *in* (4) *is absolutely continuous.*

Proof. To keep notation simple, K_c is assumed to have range in R^1—the proof for vector-valued functions is identical, but it is necessary to introduce extra subscripts. Let $g(x, u)$ denote the first three terms in (4) and let $F = \{(x, u): x \in \mathscr{A}(R^n),\ x(0) = x_0,\ u \in \mathscr{C}(R^m)\}$. As shown after (4), $\hat{c} = \inf\{g(x, u) + [z, K_c(u)]: (x, u) \in F\}$, and there exists a sequence $(x^k, u^k) \in F$ such that $g(x^k, u^k) \to \hat{c}$ and $K_c(u^k(t), t) \leqq 0$ for $t \in [0, 1]$. Also recall that z was nondecreasing.

Rudin [8, p. 166] proves that $z = r + s$, where $r \in \mathscr{A}$, $s \in \mathscr{BV}$, $\dot{s} = 0$ a.e., s is nondecreasing and $s(0) = 0$. We now suppose that $s \neq 0$ or equivalently $s(1) > 0$, and show that this leads to a contradiction. As noted above, it is possible to find $(x, u) \in F$ such that $K_c(u(t), t) \leqq 0$ for $t \in [0, 1]$, and $g(x, u) < \hat{c} + |a|s(1)/8$ where "a" was given in (SL). Since s is nondecreasing and $s(0) = 0$, the total variation of s is $s(1)$. First a summary of the proof is given.

Since $\dot{s} = 0$ a.e., it will be shown that a closed set E can be constructed that is a union of a finite number of intervals with the variation of s concentrated on E, and $\mu(E)$, the measure of E, is very small. Then a function $v \in \mathscr{C}(R^m)$ is constructed that satisfies $\|v\| \leqq \|\bar{u}\| + \|u\|$, $K_c(v(t), t) \leqq 0$ for $t \in [0, 1]$, and v agrees with $\bar{u}$, the interior control given in (SL), on E and agrees with u just outside of E. Thus

$\int_E K_c(v(t), t)\, ds(t) = \int_E K_c(\bar{u}(t), t)\, ds(t) < a \int_E ds(t) < (a/2)s(1)$ where the last inequality follows since E captures almost all the variation of s. Also $|\int_{\bar{E}} K_c(v(t), t)\, dr(t)| < |a|s(1)/8$ since $\mu(E)$ is chosen small, and $\int_{\overline{E^c}} K_c(v(t), t)\, dz(t) \leqq 0$ since z is nondecreasing and $K_c(v(t), t) \leqq 0$. Combining these inequalities, we see that $[z, K_c(v)] \leqq s(1)(a/2 + |a|/8) = 3as(1)/8$. If $\mu(E)$ is chosen small enough so that $g(x, v) < \hat{c} + |a|s(1)/4$, then $g(x, v) + [z, K_c(v)] \leqq \hat{c} + |a|s(1)/4 + as(1)3/8 = \hat{c} + as(1)/8$. Since $a, -s(1) < 0$, this contradicts the optimality of $\hat{c}$; hence $s = 0$ and $z = r \in \mathcal{A}$.

Now E and v will be constructed. Begin by choosing a closed set $H \subset [0, 1]$ with $\dot{s} = 0$ on H and $\mu(H^c) < \varepsilon$. For each $h \in H$, construct an open ball D^h of radius $2r^h$ where r^h is choosen sufficiently small so that $|s(t) - s(T)| \leqq \varepsilon|t - T|$ whenever $t, T \in \bar{D}^h$. Since $\dot{s}(h) = 0$, this construction is possible. Let B^h be the open ball centered at h of radius r^h. Since H is compact, a finite subcover of these balls $\{B_j\}$ can be chosen of radii $\{r_j\}$. Define $B = \bigcup B_j$ and $D = \bigcup D_j$; since s is monotone and $|s(t) - s(T)| \leqq \varepsilon|t - T|$ whenever $t, T \in \bar{D}_j$ for some j, then the total variation of s on $\bar{D}$ is at most ε, and hence the variation of s on D^c is at least $s(1) - \varepsilon$. Also since $H \subset B \subset D$, $\mu(D^c) \leqq \mu(B^c) \leqq \mu(H^c) < \varepsilon$. By Urysohn's lemma, there exists $\xi \in \mathscr{C}(R^1)$ satisfying $D^c < \xi < \bar{B}^c$. Defining $v = (1 - \xi)u + \xi\bar{u}$, we see that $v = u$ on B, $v = \bar{u}$ on D^c, $\|v\| \leqq \|u\| + \|\bar{u}\|$; also since $K_c(\cdot, t)$ is convex, $K_c(u(t), t) \leqq 0$ and $K_c(\bar{u}(t), t) \leqq 0$, then $K_c(v(t), t) \leqq 0$. Choosing $E = D^c$ and returning to the summary above, we notice that for ε sufficiently small, all the statements in the summary hold. □

LEMMA A.4. *Suppose* (C), (SL) *and inequality* (13) *hold. Then* $q \in \mathcal{A}(R^n)$.

Proof. Again to keep notation simple, assume q is scalar-valued—the argument below can be applied to each component of q separately to treat vector-valued functions. Since $q = G^T v - p$ on $(0, 1)$ and G is absolutely continuous while v and p lie in $\mathscr{BV}$, then q is continuous from the left on $[0, 1)$ (see the definition of $\mathscr{BV}$) and by Lemma 2, q is continuous from the left on $[0, 1]$ since $q(1^-) = 0 = q(1)$. As in Lemma A.3, we can express $q = r + s$, where $r \in \mathcal{A}$, $s \in \mathscr{BV}$, $s(0) = 0$, and $\dot{s} = 0$ almost everywhere. Let us suppose that $s(t) \neq 0$ for some $t \in [0, 1)$—it is shown that (13) is violated and hence $s = 0$.

Since s is continuous from the left, then the total variation of s on $[0, t]$ is a continuous function of t from the left, and it is possible to choose $t' < t$ such that the variation of s on $[t', t]$ is less than δ.

Using the construction of Lemma A.3 on $[0, t']$, one generates sets $B \subset D \subset [0, t']$ such that the variation of s on $\bar{D}$ is at most ε and $\mu([0, t'] - B) < \varepsilon$. (Since the construction of Lemma A.3 was only valid for a monotone function, this last step actually requires that we first express $s = s_1 + s_2$, where s_1 and $-s_2$ are both nondecreasing (see Natanson [6]) and $\dot{s}_1 = \dot{s}_2 = 0$ a.e. (see Rudin [8, p. 166]). Then using Lemma A.3, sets D_1 and D_2 are constructed that capture only $\varepsilon/4$ of the variation of s_1 and s_2, respectively, and that satisfy $\mu([0, t'] - D_1)$, $\mu([0, t'] - D_2) < \varepsilon/4$. Then define $D = D_1 \cap D_2$.

Now choose $\varepsilon < |t - t'|$ and define J_ρ to be an open ball centered at t' of diameter ρ. Again construct $\xi \in \mathscr{C}(R^1)$ satisfying $([0, t'] - D) < \xi < ([0, t'] - B) \cup J_\varepsilon$ and define $x_N = (1 - \xi)\bar{x} + \xi N$, where $\bar{x}$ was given in (SL) and $N \in R^1$ with sgn $(N) = -\text{sgn}\,(s(t))$. (We had to introduce the set J_ε since $[0, t'] - \bar{B}$ is not an open set on [0, 1] as required by Urysohn's lemma.)

Since $x_N = \bar{x}$ on $[t, 1]$,

$$\begin{aligned}\int_0^1 x_N(\sigma)\,dq(\sigma) &= \int_0^t x_N(\sigma)\,ds(\sigma) + \int_t^1 \bar{x}(\sigma)\,ds(\sigma) + \int_0^1 x_N(\sigma)\,dr(\sigma)\\ &\leqq Ns(t) + 1 + \|\bar{x}\|(TV(s) + TV(r)),\end{aligned} \tag{A.3}$$

where $TV(s)$ is the total variation of s on $[0, 1)$ and the last inequality follows from the following relations whenever ε and δ are chosen sufficiently small:

$$\begin{aligned}\int_0^{t'} x_N(\sigma)\,ds(\sigma) &= \int_{\bar{D}} x_N(\sigma)\,ds(\sigma) + \int_{[0,\,t']-D} x_N(\sigma)\,ds(\sigma)\\ &= N\int_0^{t'} ds(\sigma) + \int_{\bar{D}} (x_N(\sigma) - N)\,ds(\sigma)\\ &\leqq Ns(t') + \varepsilon(\|\bar{x}\| + |N|)\\ &\leqq -|N|(|s(t)| - \delta) + \varepsilon(\|\bar{x}\| + |N|),\end{aligned}$$

$$\int_{t'}^{t} x_N(\sigma)\,ds(\sigma) \leqq \delta(|N| + \|\bar{x}\|),$$

If $g(x)$ denotes the first three terms of (13) evaluated at $u = u^*$, then for ε and δ sufficiently small, $g(x_N)$ is close to $g(\bar{x})$. However, this combined with the inequality (A.3) and the fact that $Ns(t) < 0$ violates (13) for N large; i.e., $\inf\{g(x) + \int_0^1 x(t)^T\,dq(t) : x \in \mathscr{A}(R^n),\ x(0) = x_0\}$ is no longer finite. Hence $s = 0$ and q is absolutely continuous. □

REFERENCES

[1] W. W. HAGER, *The Ritz–Trefftz method for optimal control problems with state and control constraints*, SIAM J. Numer. Anal., 12 (1975), pp. 854–867.

[2] M. R. HESTENES, *Calculus of Variations and Optimal Control Theory*, John Wiley, New York, 1966.

[3] J. L. LIONS, *Contrôle Optimal des Systèmes Gouvernés par des Équations aux Derivées Partielles*, Dunod, Paris, 1968.

[4] N. DUNFORD AND T. SCHWARTZ, *Linear Operators*, Interscience, New York, 1963.

[5] L. W. NEUSTADT, *A general theory of extremals*, J. Comput. System Sci., 3 (1969), pp. 57–59.

[6] I. P. NATANSON, *The Theory of Functions of a Real Variable*, Frederick Ungar, New York, 1964.

[7] R. T. ROCKAFELLAR, *State constraints in convex control problems of Bolza*, this Journal, 10 (1972), pp. 691–715.

[8] W. RUDIN, *Real and Complex Analysis*, McGraw-Hill, New York, 1966.

[9] E. B. LEE AND L. MARKUS, *Foundations of Optimal Control Theory*, John Wiley, New York, 1967.

[10] J. BARROS-NETO, *An Introduction to the Theory of Distributions*, Marcel Dekker, New York, 1973.

[11] W. W. HAGER, *Quadratic program stability and optimal control regularity*, to appear.