SIAM J. CONTROL AND OPTIMIZATION
Vol. 15, No. 2, February 1977

THE HIGH ORDER MAXIMAL PRINCIPLE AND ITS APPLICATION TO SINGULAR EXTREMALS*

ARTHUR J. KRENER†

Abstract. The high order maximal principle (HMP) which was announced in [11] is a generalization of the familiar Pontryagin maximal principle. By using the higher derivatives of a large class of control variations, one is able to construct new necessary conditions for optimal control problems with or without terminal constraints. In particular, we show how the HMP can be used to prove the generalized Legendre–Clebsch condition of Kelley, Kopp, Moyer and Goh. The principle advantage of this derivation is that, unlike previous ones, it remains valid even when there are terminal constraints.

1. Introduction. Although we are interested in high order necessary conditions for optimal control problems, let us first consider the following nonlinear programming problem. Minimize the smooth function $y_0(x)$ subject to the smooth constraints $y_i(x)=0$ for $i=1,\cdots,m$ and $x\in\mathcal{A}\subseteq\mathbb{R}^n$. The set $\mathcal{A}$ is not explicitly described, instead, given $x^e\in\mathcal{A}$ we assume there are ways of generating smooth curves $s\mapsto x(s)\in\mathcal{A}$ for $s\in[0,\varepsilon)$ such that $x(0)=x^e$. To develop first order necessary conditions for this problem we adjoin the constraints y_i to y_0 via Lagrange multipliers $\nu_0,\nu_1,\cdots,\nu_m$, where ν_0 is normalized to be nonpositive. If x^e is a minimum, then every curve $x(s)$ as above generates a necessary condition

$$\frac{d}{ds}\sum_{i=0}^{m}\nu_i y_i(x(0))=\sum_{i=0}^{m}\nu_i\frac{\partial}{\partial x}y_i(x^e)\frac{d}{ds}x(0)\leqq 0. \tag{1.1}$$

The use of the Lagrange multipliers requires some assumption of local convexity on the set $\{x: y_i(x)=0,\ i=1,\cdots,m\}\cap\mathcal{A}$ around x^e. Since $\mathcal{A}$ is not explicitly given, this cannot be verified. Instead we assume the following: the gradients of the functions $y_0,\cdots,y_m$ are linearly independent at x^e and whenever $x^1(s)$ and $x^2(s)$ are used to develop necessary conditions via (1.1), for any $0\leqq \mu\leqq 1$ there exists a curve $x^3(s)\in\mathcal{A}$ such that $x^3(0)=x^e$ and

$$\frac{d}{ds}x^3(0)=\mu\frac{d}{ds}x^1(0)+(1-\mu)\frac{d}{ds}x^2(0). \tag{1.2}$$

As we shall see later, this form of convexity suffices to justify the multipliers. Of course if $m=0$, no convexity assumption or multipliers are needed and ν_0 can be set to be -1.

The goal of any collection of necessary conditions is to isolate a hopefully unique candidate for the minimum. Additional conditions may be required to narrow the field of possibilities and to distinguish between potential maxima and minima. If a collection of necessary conditions of the form (1.1) does not

* Received by the editors October 14, 1975, and in revised form February 10, 1976.

† Department of Mathematics, University of California, Davis, California 95616. This research was supported in part by the U.S. Office of Naval Research under the Joint Services Electronics Program Contract N00014-75-C-0648, while the author was a research fellow at the Division of Engineering and Applied Physics, Harvard University, Cambridge, Massachusetts, 1974–75, and also by the National Science Foundation under Grant MPS75-05248.

completely accomplish this task, one can look for additional curves $x(s)$ or obtain higher order conditions by differentiating (1.1) further.

If there are no y_i constraints ($m=0$), then it is clear that the first nonzero derivative of $\nu_0 y_0(x(s))$ must be negative for x^e to be a minimum. In general this involves higher order partial derivatives of $y_0(x)$. For example, if (1.1) is assumed to be zero, then the second derivative test is

$$\nu_0 \frac{d^2}{ds^2} y_0(x(0)) = \nu_0 \left(\frac{d}{ds} x(0)\right)^T \frac{\partial^2}{\partial x^2} y_0(x^e) \left(\frac{d}{ds} x(0)\right) + \nu_0 \frac{\partial}{\partial x} y_0(x^e) \frac{d^2}{ds^2} x(0) \leqq 0. \tag{1.3}$$

Suppose $x^e \in$ interior $\mathscr{A} \subseteq \mathbb{R}^n$. Then (1.1) implies that $(\partial/\partial x) y_0(x^e) = 0$ and so (1.3) reduces to the familiar condition where the Hessian $\nu_0(\partial^2/\partial x^2) y_0(x^e)$ is negative semidefinite at a minimum. On the other hand, if for some $x(s)$, $(d/ds)x(0) = 0$, then (1.1) is trivially satisfied and (1.3) yields a condition which involves only the gradient of y_0. The same condition can be obtained from (1.1) by reparametrizing $x(s)$ as $x(s^{1/2})$.

If there are terminal constraints, then second order conditions similar to (1.3) can be developed with some difficulty, since the use of the Lagrange multiplier must be justified. For higher derivatives, this justification is so difficult as to make the resulting necessary conditions of little practical value. The difficulties arise because, in general, these conditions involve second and higher order partial derivatives of $y_0, y_1, \cdots, y_m$. As was seen above, there is an exception to that; if the first $h-1$ derivatives of $x(s)$ are zero at $s=0$, then

$$\frac{d^h}{ds^h} \sum_{i=0}^{m} \nu_i y_i(x(0)) = \sum_{i=0}^{m} \nu_i \frac{\partial}{\partial x} y_i(x^e) \frac{d^h}{ds^h} x(0) \leqq 0 \tag{1.4}$$

involves only the first partial derivatives of $y_0, \cdots, y_m$. In this case, justifying the Lagrange multiplier requires only a convexity assumption for higher derivatives similar to (1.2). It is this type of necessary condition which we consider in this paper.

Now we turn to optimal control problems which generate nonlinear programming problems of the type we have been considering. Suppose we wish to minimize $y_0(x(t^e))$ subject to $\dot{x} = f(x(t), u(t))$, $x(t^0) = x^0$, $y_i(x(t^e)) = 0$, $i = 1, \cdots, m$, and $u(t) \in \Omega$ for $t \in [t^0, t^e]$. Let $\mathscr{A}$ denote the set of points accessible from x^0 using admissible controls. Suppose a control $u(t)$ and trajectory $x(t)$ defined on $[t^0, t^e]$ is a candidate for an optimal solution. We can generate curves lying in $\mathscr{A}$ by considering the locus of endpoints $x(t^e; s)$ of a family of trajectories $x(t; s)$ generated by controls $u(t; s)$ which are variations of $x(t)$ and $u(t)$ depending on the parameter s. The controls $u(t; s)$ are obtained by replacing $u(t)$ by some other control $v(t)$ for $t \in [t^1 - s, t^1]$ where $t^1 \in (t^0, t^e)$. The reference control and trajectory are obtained when $s = 0$. In this way, using (1.1), one develops the usual linear necessary conditions, i.e., the Pontryagin maximum principle (PMP), which is most conveniently expressed in a Hamiltonian format.

It frequently happens in nonlinear control problems that the set of first derivatives of the curves obtained by the above procedure does not fully represent all the degrees of freedom within the set $\mathscr{A}$ of accessible points around the reference endpoint $x(t^e)$. Such controls and trajectories are called *singular* (in the sense of the PMP as opposed to the classical definition in the calculus of variations). For this reason the PMP can prove to be inadequate in determining either a unique candidate or distinguishing between minimizing and maximizing trajectories. (See [2].)

The high order maximum principle [HMP) is an attempt to overcome these difficulties. More complicated control variations are used which have the property that lower order derivatives of $x(t^e; s)$ are zero and the first nonzero derivatives lie in directions within $\mathscr{A}$ which were not available as first derivatives. Since the lower derivatives are zero and a convexity assumption for higher derivatives similar to (1.2) is satisfied, equation (1.4) can be applied to obtain new necessary conditions which can also be expressed in terms of the Hamiltonian.

The organization of the rest of the paper is as follows. The statement of the HMP is found in § 2 and the proof in § 3. Then the HMP is used to develop linear and quadratic necessary conditions for singular extremals. Scalar controls are treated in §§ 4 and 5, and in § 6, vector controls are treated. (These conditions are called linear and quadratic not because they are linear or quadratic with respect to the parameter s mentioned above, but rather because they are linear or quadratic with respect to the L^1 norm of the control variation. We elaborate on this later.)

The linear conditions are those implied by the PMP. The quadratic conditions reduce to the generalized Legendre–Clebsch (GLC) of Kelley, Kopp and Moyer [8] (scalar control) and Goh [4] (vector controls) when the problem in question is normal or there are no terminal constraints. Using the HMP we can extend the GLC to problems which do not satisfy these assumptions.

We wish to emphasize that these are not the only applications of the HMP, rather, the HMP is a very powerful tool for constructing necessary conditions, the simplest of which are the ones mentioned above. We hope that by studying this paper the reader will be able to construct new necessary conditions in an ad hoc fashion which are appropriate to the problem of interest.

2. The high order maximal principle. Consider a system whose dynamics are given by

$$\dot{x} = f(x, u) \tag{2.1}$$

subject to $x(t^0) = x^0$ and $u(t) \in \Omega$, where $x = (x_0, x_1, \cdots, x_n)$ with $x_0 = t$, $u = (u_1, \cdots, u_l)$, f a C^∞-function of x and u, Ω some subset of $\mathbb{R}^l$. The state variables x are local coordinates on an $(n+1)$-dimensional C^∞-manifold M. However we proceed as if they are globally defined and leave to the reader the task of "patching things together", i.e., supplying the intrinsic meaning for all of the objects described in a coordinate-dependent fashion.

The problem is to find a piecewise C^∞-control $u(t) \in \Omega$ for $t \in [t^0, t^e]$ which generates a trajectory $x(t)$ satisfying the boundary conditions

$$x(t^0) = x^0 \quad \text{and} \quad y_i(x(t^e)) = 0, \qquad i = 1, \cdots, m, \tag{2.2}$$

which minimizes

$$y_0(x(t^e)). \tag{2.3}$$

The functions $y_0, y_1, \cdots, y_m$ are assumed to be C^∞ and linearly independent everywhere of interest. Since time is a state variable, (2.1) could be time-dependent and the functions $y_0, \cdots, y_m$ could also depend on time. Control problems where the integral of a Lagrangian are to be minimized can easily be converted to the above format by the addition of another state variable.

The assumption of infinite differentiability is not required, it is only invoked to avoid counting the degree of differentiability needed in a particular argument. Piecewise differentiability means left and right limits always exist and there are only a finite number of jumps in any compact interval. Throughout the paper we assume that the controls being considered are C^∞ at the times in question. At other times, similar results can be deduced by restricting to left or right limits and by continuity. Since the details are tedious, we choose the convenient expedient of leaving them to the reader.

Corresponding to each *admissible control*, $u^i(t) \in \Omega$, is an *admissible vector field*

$$f^i(x) = f(x, u^i(x_0))$$

which generates an *admissible flow* $\gamma^i(s)x$ defined as the family of integral curves of the differential equation

$$\frac{d}{ds}\gamma^i(s)x = f^i(\gamma^i(s)x)$$

satisfying the initial conditions

$$\gamma^i(0)x = x.$$

Suppose the reference trajectory $x(t) = \gamma^0(t-t^0)x^0$ is generated by the control $u^0(t)$ for $t \in [t^0, t^e]$. Then a standard proof of the PMP is to replace the reference control by another control $u^1(t)$ for $t \in [t^1 - s, t^1]$ where $t^1 \in (t^0, t^e)$. The result is a family of trajectories $x(t; s)$ indexed by small $s \geqq 0$ whose locus of endpoints is given by

$$\begin{aligned} x(t^e; s) &= \gamma^0(t^e - t^1)\gamma^1(s)\gamma^0(t^1 - t^0 - s)x^0 \\ &= \gamma^0(t^e - t^1)\gamma^1(s)\gamma^0(-s)x^1, \end{aligned}$$

where $x^1 = x(t^1)$. If we define $\alpha(s)x = \gamma^1(s)\gamma^0(-s)x$, then this can be written as

$$\gamma^0(t^e - t^1)\alpha(s)x^1.$$

For this reason, we call the map $\alpha(s)x$ a *control variation* to u^0 before x.

Alternately, $u^0(t)$ could be replaced by $u^1(t)$ on the interval $[t^1, t^1 + s]$ resulting in a locus of endpoints

$$\begin{aligned} x(t^e; s) &= \gamma^0(t^e - t^1 - s)\gamma^1(s)x^1 \\ &= \gamma^0(t^e - t^1)\gamma^0(-s)\gamma^1(s)x^1. \end{aligned}$$

This time we have a control variation $\alpha(s)x = \gamma^0(-s)\gamma^1(s)x$ to u^0 after x. Various combinations of the above are possible, for example, $\alpha(s)x = \gamma^0(-s/2)\gamma^1(s)\gamma^0(-s/2)x$, a control variation to u^0 at x. As we shall see in § 3, if

$u^0(t)$ and $u^1(t)$ are smooth at t^1, then all of the above yield the same necessary conditions.

The important point about these variations $\alpha(s)x$ is that when they are inserted into a trajectory generated by the control $u^0(t)$, the result is a family of admissible trajectories indexed by small $s \geqq 0$ whose locus of endpoints is a smooth function of s. With this in mind we define a *control variation* $\alpha(s)x$ to $u^0(t)$ at x as being of the following form:

$$\alpha(s)x = \gamma^0(q_2(s))\gamma^k(p_k(s)) \cdots \gamma^1(p_1(s))\gamma^0(q_1(s))x, \tag{2.4}$$

where $\gamma^0, \gamma^1, \cdots, \gamma^k$ are the flows of admissible controls $u^0(t), u^1(t), \cdots, u^k(t)$ and $q_i(s)$ and $p_i(s)$ are polynomials in s satisfying $q_i(0) = p_i(0) = 0$ and $p_i(s) \geqq 0$ for small $s \geqq 0$. This is similar to the bundle variation of Gabasov and Kirillova [2]. The reader should note that

$$x(t^e; s) = \gamma^0(t^e - t^1)\alpha(s)\gamma^0(t^1 - t^0)x^0$$

is the locus of endpoints of a family of admissible trajectory for small $s \geqq 0$, and hence a curve in $\mathcal{A}$. Moreover, $x(t^e; s)$ is a smooth function of s and $x(t^e; 0)$ is the endpoint of the reference trajectory. Notice that if $q_1(s) + q_2(s) + \sum p_i(s) \neq 0$, then the control variation changes the terminal time, t^e. In particular, the variations $\alpha(s)x = \gamma^0(\pm s)x$ lengthen or shorten the reference trajectory.

A control variation $\alpha(s)x$ is said to be of *order* h at $x^1 = x(t^1)$ if there exists an $\varepsilon > 0$ such that

$$\frac{d^j}{ds^j}\alpha(0)x(t) = 0 \tag{2.5}$$

for $j = 1, \cdots, h-1$ and $|t - t^1| < \varepsilon$. In particular, for $h = 1$, there are no lower derivatives for which (2.5) must hold and so every variation is of order at least one. A control variation of order h is a fortiori of order $1, \cdots, h-1$. Because the earlier derivatives are zero, it is the h derivative of $x(t^e; s)$ which supplies the necessary condition via (1.4). As we show in the next section, it is necessary to require (2.5) to hold in a time interval around t^1 so that the convexity assumption for higher derivatives holds and the use of multipliers can be justified.

The high order maximal principle (HMP). Let $u^0(t)$ be an admissible control generating the trajectory $x(t) = \gamma^0(t - t^0)x^0$ for $t \in [t^0, t^e]$. If $u^0(t)$ minimizes $y_0(x(t^e))$ subject to the boundary condition $y_i(x(t^e)) = 0$ for $i = 1, \cdots, m$, then there exists a nontrivial adjoint variable $\lambda(t) = (\lambda_0(t), \cdots, \lambda_n(t))$ defined for $t \in [t^0, t^e]$ and satisfying

$$\dot\lambda(t) = -\lambda(t)\frac{\partial}{\partial x}f(x(t), u^0(t)), \tag{2.6}$$

$$\lambda(t^e) = \sum \nu_i \frac{\partial}{\partial x} y_i(x(t^e)), \quad \text{where } \nu_0 \leqq 0, \tag{2.7}$$

$$\lambda(t)f(x(t), u) \leqq \lambda(t)f(x(t), u^0(t)) = 0 \quad \forall\, u \in \Omega, \tag{2.8}$$

and for every control variation $\alpha(s)x$ of order h at $x(t)$,

$$\lambda(t)\frac{d^h}{ds^h}\alpha(0)x(t) \leqq 0. \tag{2.9}$$

Conditions (2.6), (2.7) and (2.8) are the familiar PMP and (2.1), (2.6) and (2.8) can be conveniently expressed in terms of the Hamiltonian, $H(\lambda, x, u) = \lambda f(x, u)$, as Hamilton's differential equation

$$\dot{x} = \frac{\partial}{\partial \lambda} H(\lambda(t), x(t), u^0(t)), \tag{2.10}$$

$$\dot{\lambda} = -\frac{\partial}{\partial x} H(\lambda(t), x(t), u^0(t)) \tag{2.11}$$

and the Pontryagin–Weierstrass condition,

$$0 = H(\lambda(t), x(t), u^0(t)) = \max_{u \in \Omega} H(\lambda(t), x(t), u). \tag{2.12}$$

(Equations (2.8) and (2.12) are zero because $x_0 = t$.) A $u^0(t)$ and $x(t)$ for which there exists a $\lambda(t)$ satisfying (2.6) and (2.8) are called an *extremal control and extremal trajectory*. If $u^0(t) \in$ interior Ω, then (2.12) implies

$$\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) = 0 \tag{2.13}$$

and the Legendre–Clebsch condition

$$\frac{\partial^2}{\partial u^2} H(\lambda(t), x(t), u^0(t)) \leqq 0. \tag{2.14}$$

The new condition (2.9) is a generalization of the Pontryagin–Weierstrass condition to control variations of higher order. It in turn leads to a generalization of the Legendre–Clebsch condition for extremal trajectories which are singular in the classical sense (i.e., (2.14) not of full rank). This is demonstrated in §§ 5 and 6.

We have fixed the initial point $x(t^0) = x^0$ but there is a straightforward extension of the HMP to problems where the initial point is only partially constrained. In this case (2.6)–(2.9) still hold and, in addition, $\lambda(t^0)$ must satisfy a transversality condition similar to (2.7) as in the PMP.

When applying the HMP it is highly desirable to choose a minimal realization of the problem under consideration where $y(x) = (y_0(x), \cdots, y_m(x))$ is considered as the output map. (For the theory of minimal realizations of nonlinear systems, see Sussmann [16].) The reason for this is that the less state dimensions there are, the easier it is to find higher order control variations satisfying (2.5), and so the more necessary conditions result. An example of just this point is given in § 5.

3. Proof of the HMP. We start by noting that the order of a control variation can easily be shifted upward.

LEMMA 3.1. *Suppose $\alpha^1(s)x$ is a control variation of order h at x^1. Then for any integer k there exists a control variation $\alpha^2(s)x$ of order $h \cdot k$ at x^1 whose $h \cdot k$ derivative is a positive multiple of the h derivative of $\alpha^1(s)x$ and hence yields the same necessary condition when used in* (2.9) *of the HMP.*

Proof. Define $\alpha^2(s)x = \alpha^1(s^k)x$. It is straightforward to verify that $\alpha^2(s)x$ is a control variation as in (2.4) and of order $h \cdot k$ with the $h \cdot k$ derivative as described.

To compute higher derivatives of control variations, it is convenient to let them operate on smooth functions as a partial differential operator.

LEMMA 3.2. *A necessary and sufficient condition for $\alpha(s)x$ to be a control variation of order h at x^1 is that for some $\varepsilon > 0$ and for every C^∞ real-valued function $\varphi(x)$ defined in some neighborhood of x^1,*

$$\frac{d^j}{ds^j}\varphi(\alpha(0)x(t)) = 0$$

for $j = 1, \cdots, h-1$ and $|t-t^1| < \varepsilon$. Moreover if $\alpha(s)x$ is of order h at x^1,

$$\frac{d^h}{ds^h}\varphi(\alpha(0)x^1) = \frac{\partial}{\partial x}\varphi(x^1)\frac{d^h}{ds^h}\alpha(0)x^1.$$

Proof. The proof is straightforward.

We next show that if all the controls involved are C^∞ at t^1, then it does not matter whether a variation is made before, after or at $x^1 = x(t^1)$.

LEMMA 3.3. *Let $\alpha^1(s)x$ be a control variation to $u^0(t)$ of order h at x^1 and $q(s)$ be a polynomial in s such that $q(0) = 0$. Define a new control variation $\alpha^2(s)x = \gamma^0(q(s))\alpha^1(s)\gamma^0(-q(s))x$. Then α^2 is also of order h at x^1 and furthermore yields the same necessary condition in* (2.9) *for*

$$\frac{d^h}{ds^h}\alpha^2(0)x^1 = \frac{d^h}{ds^h}\alpha^1(0)x^1.$$

Proof. Consider $\alpha^2(s)x$ as a function of four variables,

$$\alpha^2(s_1, s_2, s_3)x = \gamma^0(q(s_3))\alpha^1(s_2)\gamma^0(-q(s_1))x,$$

where $s_1 = s_2 = s_3 = s$. Then for any C^∞-function, by the chain rule,

$$\frac{d^l}{ds^l}\varphi(\alpha^2(0)x) = \sum\binom{l}{i, j, k}\frac{\partial^i}{\partial s_1^i}\frac{\partial^j}{\partial s_2^j}\frac{\partial^k}{\partial s_3^k}\varphi(\alpha^2(0, 0, 0)x), \tag{3.1}$$

where the sum is over all $i, j, k \geqq 0$, $i + j + k = l$. For any k define a C^∞-function,

$$\psi(x) = \frac{\partial^k}{\partial s_3^k}\varphi(\gamma^0(q(0))x).$$

For small s_1 there exists a $t(s_1)$ near t^1 such that

$$\gamma^0(-q(s_1))x^1 = x(t(s_1)).$$

Since $\alpha^1(s)x$ is of order h at x^1, for $1 \leqq j \leqq h-1$ and small s_1,

$$\frac{\partial^j}{\partial s_2^j}\frac{\partial^k}{\partial s_3^k}\varphi(\alpha^2(s_1, 0, 0)) = \frac{\partial^j}{\partial s_2^j}\psi(\alpha^1(0)\gamma^0(-q(s_1))x^1 = \frac{\partial^j}{\partial s_2^j}\psi(\alpha^1(0)x(t(s_1))) = 0.$$

So for $1 \leqq l \leqq h-1$, equation (3.1) becomes

$$\frac{d^l}{ds^l}\varphi(\alpha^2(0)x^1) = \sum_{i=0}^{l}\binom{l}{i}\frac{\partial^i}{\partial s_1^i}\frac{\partial^{l-i}}{\partial s_3^{l-i}}\varphi(\alpha(0,0,0)x^1)$$

$$= \frac{d^l}{ds^l}\varphi(\gamma^0(q(0))\gamma^0(-q(0))x^1)$$

$$= \frac{d^l}{ds^l}\varphi(x^1) = 0,$$

since $\gamma^0(q(s))\gamma^0(-q(s))x^1 = \gamma^0(q(s)-q(s))x^1 = \gamma^0(0)x^1 = x^1$. The same arguments can be repeated at each $x(t)$, $|t-t^1| < \varepsilon$ to show $\alpha^2(s)x$ satisfies (2.5) in an interval around t^1.

Similarly evaluating (3.1) for $l = h$, we have

$$\frac{d^h}{ds^h}\varphi(\alpha^2(0)x^1) = \frac{d^h}{ds^h}\varphi(\alpha^1(0)x^1) + \frac{d^h}{ds^h}\varphi(\gamma^0(q(0))\gamma^0(-q(0))x^1)$$

$$= \frac{d^h}{ds^h}\varphi(\alpha^1(0)x^1).$$

Q.E.D.

If the control $u^0(t)$ is not continuous at t^1, then the trajectory has a corner at x^1 and the effect of control variations on either side are different. By comparing these differences one can deduce various corner conditions for optimality, but this is a topic we shall not pursue any further. We refer the interested reader to Kelley, Kopp and Moyer [8], Gabasov and Kirillova [2], McDanell and Powers [12] and Maurer [13].

The next two lemmas are crucial to the HMP because they show that for higher order control variations satisfying (2.5), one can "add" them and, in particular, form convex combinations as required by the use of Lagrange multipliers. First we deal with control variations made at the same point of the trajectory.

LEMMA 3.4. *Suppose $\alpha^1(s)x, \cdots, \alpha^r(s)x$ are control variations to $u^0(t)$ at $x^1 = x(t^1)$ of order $h_1, \cdots, h_r$ respectively. Let $c = (c_1, \cdots, c_r)$ be a vector of nonnegative real numbers. Then there exists a family of control variations $\alpha(s; c)x$ of order h (= the least common multiple $\{h_i\}$) such that*

$$\frac{d^h}{ds^h}\alpha(0; c)x^1 = \sum_{i=1}^{r} c_i \frac{d^{h_i}}{ds^{h_i}}\alpha^i(0)x^1.$$

Moreover $\alpha(s, c)x$ is continuous in c for small $s \geqq 0$.

Proof. For notational simplicity, assume $r = 2$; the general case follows by a similar argument. Using Lemma 3.1, we can assume that $h = h_1 = h_2$, and using Lemma 3.3, that both variations are made before x^1, for example,

$$\alpha^1(s)x = \gamma^k(p_k(s)) \cdots \gamma^1(p_1(s))\gamma^0(q(s))x.$$

Define a family of new variations

$$\alpha(s; c)x = \gamma^k(p_k(c_1^{1/h}s)) \cdots \gamma^1(p_1(c_1^{1/h}s))\alpha^2(c_2^{1/h}s)\gamma^0(q(c_1^{1/h}s))x.$$

Introduce parameters $s_1 = s_2 = s_3 = s$ into $\alpha(s; c)x$ as in the proof of Lemma 3.3. Then for any C^∞-function φ,

$$\frac{d^l}{ds^l}\varphi(\alpha(0; c)x^1) = \sum \binom{l}{i, j, k} c_1^{i/h} c_2^{j/h} c_1^{k/h} \frac{\partial}{\partial s_1^i} \frac{\partial}{\partial s_2^j} \frac{\partial^k}{\partial s_3^k} \varphi(\alpha(0, 0, 0; c)x^1).$$

If $1 \leqq l \leqq h-1$, this reduces as before to

$$\begin{aligned}\frac{d^l}{ds^l}\varphi(\alpha(0; c)x^1) &= \sum_{i=0}^{l} \binom{l}{i} c_1^l \frac{\partial^i}{\partial s_1^i} \frac{\partial^{l-i}}{\partial s_3^{l-i}} \varphi(\alpha(0, 0, 0; c)x^1)\\ &= c_1^l \frac{d^l}{ds^l}\varphi(\alpha^1(0)x^1) = 0,\end{aligned}$$

since α^1 is of order h at x^1.

For $l = h$ we have

$$\frac{d^h}{ds^h}\varphi(\alpha(0; c)x^1) = c_1 \frac{d^h}{ds^h}\varphi(\alpha^1(0)x^1) + c_2 \frac{d^h}{ds^h}\varphi(\alpha^2(0)x^1). \qquad \text{Q.E.D.}$$

If a control variation $\alpha(s)x$ of order h is made at x^1, then the result is a family of trajectories whose locus of endpoints is given by

$$x(t^e; s) = \gamma^0(t^e - t^1)\alpha(s)x^1.$$

The first $h-1$ derivatives of $x(t^e; s)$ are zero and h derivative is given by

$$\frac{d^h}{ds^h} x(t^e; 0) = \left(\frac{\partial}{\partial x}\gamma^0(t^e - t^1)x^1\right) \frac{d^h}{ds^h}\alpha(0)x^1. \tag{3.2}$$

This is applied to (1.4) to obtain (2.9) of the HMP, but first we must show that we can "add" the effect of control variations made at differing times.

LEMMA 3.5. *Suppose $\alpha^1(s)x, \cdots, \alpha^r(s)x$ are control variations to $u^0(t)$ at $x^1 = x(t^1), \cdots, x^r = x(t^r)$ of order $h_1, \cdots, h_r$ respectively. Let $c = (c_1, \cdots, c_r)$ be a vector of nonnegative real numbers. Then there exists a family of admissible trajectories indexed by small $s \geqq 0$ and c whose locus of endpoints is given by $x(t^e; s; c)$ such that*

$$\frac{d^j}{ds^j} x(t^e; 0; c) = 0$$

for $j = 1, \cdots, h-1$ where h = least common multiple $\{h_i\}$ and

$$\frac{d^h}{ds^h} x(t^e; 0; c) = \sum_{i=1}^{r} c_i \left(\frac{\partial}{\partial x}\gamma^0(t^e - t^i)x^i\right) \frac{d^{h_i}}{ds^{h_i}}\alpha^i(0)x^i. \tag{3.3}$$

Moreover $x(t^e; s; c)$ is continuous in c for small $s \geqq 0$.

Proof. Using Lemma 3.4, we can assume that the t^i are distinct. For simplicity, assume $r = 2$, $t^1 < t^2$ and $h_1 = h_2 = h$. The general case follows by a similar argument. Consider the family of trajectories whose locus of endpoints is given by

$$x(t^e; s; c) = \gamma^0(t^e - t^2)\alpha^2(c_2^{1/h}s)\gamma^0(t^2 - t^1)\alpha^1(c_1^{1/h}s)\gamma^0(t^1 - t^0)x^0.$$

Suppose φ is a C^∞-function at $x^e = x(t^e)$. Then using the chain rule technique,

$$\frac{d^l}{ds^l}\varphi(x(t^e; 0; c)) = \sum_{i=0}^{l} \binom{l}{i} c_1^{i/h} c_2^{(l-i)/h} \frac{\partial^i}{\partial s_1^i} \frac{\partial^{l-i}}{\partial s_2^{l-i}} \varphi(x(t^e; 0; c)).$$

Let

$$\psi(x) = \frac{\partial^{l-i}}{\partial s_2^{l-i}} \varphi(\gamma^0(t^e - t^2)\alpha^2(c_2^{1/h}0)\gamma^0(t^2 - t^1)x).$$

Then $\psi(x)$ is a C^∞-function at $x(t^1)$. Since α^1 is of order h, for $1 \leqq i < h$,

$$\frac{\partial^i}{\partial s_1^i} \psi(\alpha^1(c_1^{1/h}0)x^1) = 0.$$

Therefore if $1 \leqq l < h$, then

$$\frac{d^l}{ds^l}\varphi(x(t^e; 0; c)) = c_2^{l/h} \frac{\partial^l}{\partial s_2^l}(\varphi \circ \gamma^0(t^e - t^2))\alpha^2(c_2^{1/h}0)x^2 = 0$$

since α^2 is of order h. A similar argument proves (3.3). Q.E.D.

In light of this lemma, we define a cone K in the tangent space at $x^e = x(t^e)$ as the convex hull of all vectors of the form (3.2). This cone is a measure of the controllability at x^e available through higher order control variations made all along the reference trajectory. The completion of the proof of the HMP follows Halkin's proof of the PMP [5] using a fixed point argument. Intuitively for $u^0(t)$ to be minimal, the cone K of controllability must be separable by a hyperplane from the cone of L of directions which satisfy the boundary conditions and decrease y_0. Formally L is defined to be the cone of all tangent vectors τ at x^e such that

$$\left(\frac{\partial}{\partial x} y_0(x^e)\right)\tau \leqq 0$$

and for $i = 1, \cdots, m$,

$$\left(\frac{\partial}{\partial x} y_i(x^e)\right)\tau = 0.$$

THEOREM 3.6 (HMP). *Suppose there exists no nontrivial adjoint variable satisfying* (2.6)–(2.9). *Then $u^0(t)$ is not minimal.*

Proof. If $\lambda^e = (\lambda_0^e, \cdots, \lambda_n^e)$ defines a hyperplane separating K and L in the tangent space of x^e, i.e.,

$$\lambda^e \tau \leqq 0 \quad \forall \tau \in K,$$

$$\lambda^e \tau \geqq 0 \quad \forall \tau \in L,$$

then define

$$\lambda(t) = \lambda^e \frac{\partial}{\partial x} \gamma^0(t^e - t)x^0.$$

It is easy to verify that $\lambda(t)$ satisfies (2.6)–(2.9).

On the other hand, if no such λ^e exists, then it follows that there exists no hyperplane separating K^* and L^* where these are the cones in $\mathbb{R}^{m+1}$ defined by

$$K^* = \left\{\frac{\partial}{\partial x} y(x^e)\tau : \tau \in K\right\},$$

$$L^* = \left\{\frac{\partial}{\partial x} y(x^e)\tau : \tau \in L\right\}.$$

(Recall that $y = (y_0, \cdots, y_m)$ and the $(m+1)\times(n+1)$ matrix $\partial y/\partial x(x^e)$ is assumed to be of full rank, $m+1$.)

From the definition of L, the cone L^* is generated by the vector $(-1, 0, \cdots, 0)$, hence this vector must be in the interior of K^*. Suppose $\sigma^0, \cdots, \sigma^m$ are linearly independent vectors in K^* such that

$$(-1, 0, \cdots, 0) = \sum_{i=0}^{m} \sigma^i.$$

Let $\tau^0, \cdots, \tau^m$ be vectors in K such that

$$\sigma^i = \frac{\partial}{\partial x} y(x^e)\tau^i.$$

For some h and for each $i = 0, \cdots, m$, there is a control variation $\alpha^i(s)x$ made at some $x(t^i)$ such that (3.2) equals τ^i. These variations can be used to construct a family of admissible trajectories whose locus of endpoints is given by $x(t^e; s; c)$ as in Lemma 3.5.

The vectors $\sigma^0, \cdots, \sigma^m$ form a basis for $\mathbb{R}^{m+1}$ and we use $\|\cdot\|$ to denote the L^1 norm relative to this basis, i.e. $\|\sum d_i\sigma^i\| = \sum |d_i|$. In particular if $r \geqq 0$, $c_i \geqq 0$ and $\sum c_i = 1$, then $\|r \sum c_i\sigma^i\| = r$.

By Taylor's theorem and compactness, there exists a constant M and an $\varepsilon > 0$ such that

$$\left\| y(x(t^e; s; c)) - y(x^e) - \frac{s^h}{h!} \sum_{i=0}^{m} c_i\sigma^i \right\| \leqq Ms^{h+1} \tag{3.4}$$

for all $\{(s, c): 0 \leqq s \leqq \varepsilon,\ c_i \geqq 0,\ \sum c_i = 1\}$.

For some $\varepsilon_1 > 0$, let

$$S = \left\{ r \sum_{i=0}^{m} c_i\sigma^i : 0 \leqq r \leqq \varepsilon_1,\ c_i \geqq 0 \text{ and } \sum c_i = 1 \right\}$$

and

$$\sigma^* = (-\varepsilon_1/2, 0, \cdots, 0).$$

Clearly $\sigma^* \in$ interior $S \subseteq K^*$. Define a map $g: S \to \mathbb{R}^{m+1}$ by

$$g\left(r \sum_{i=0}^{m} c_i\sigma^i\right) = y(x(t^e; (h!r)^{1/h}; c)).$$

Then from (3.4) we see that if ε_1 is small enough, there exists a constant M_1 such that

$$\left\| g\left(r \textstyle\sum_{i=0}^{m} c_i\sigma^i\right) - \left(y(x^e) + r \displaystyle\sum_{i=0}^{m} c_i\sigma^i\right)\right\| \leqq M_1 r^{1+1/h}.$$

Let $N(\sigma^*, \delta)$ denote the closed ball of radius δ around σ^* in the norm $\|\cdot\|$. Choose δ small enough so that this neighborhood is contained in S and choose $0<\theta<1$ such that

$$M_1\theta^{1+1/h}(\delta+\varepsilon_1/2)^{1+1/h} < \theta\delta. \tag{3.5}$$

Since S is a convex set containing both 0 and $N(\sigma^*, \delta)$, it follows that it contains $N(\theta\sigma^*, \theta\delta)$. Finally define

$$g_1\left(r \sum_{i=0}^{m} c_i\sigma^1\right) = y(x^e) + r \sum_{i=0}^{m} c_i\sigma^i - g\left(r \sum_{i=0}^{m} c_i\sigma^i\right) + \theta\sigma^*.$$

Clearly g_1 is continuous and we claim that g_1 maps $N(\theta\sigma^*, \theta\delta)$ into itself. To see this, suppose $r \sum c_i\sigma^i \in N(\theta\sigma^*, \theta\delta)$. Then

$$\begin{aligned} \left\| g_1\left(r \sum_{i=0}^{m} c_i\sigma^i\right) - \theta\sigma^*\right\| &= \left\| y(x^e) + r \sum_{i=0}^{m} c_i\sigma^i - g\left(r \sum_{i=0}^{m} c_i\sigma^i\right)\right\| \\ &\leqq M_1 r^{1+1/h}. \end{aligned} \tag{3.6}$$

By the triangle inequality,

$$r = \left\| r \sum_{i=0}^{m} c_i\sigma^i\right\| \leqq \|\theta\sigma^*\| + \theta\delta = \theta(\varepsilon_1/2+\delta).$$

Putting these two inequalities together with (3.5) we obtain

$$\left\| g_1\left(r \sum_{i=0}^{m} c_i\sigma^i\right) - \theta\sigma^*\right\| < \theta\delta$$

as desired.

By the Brouwer fixed point theorem there exists an $r \sum c_i\sigma^i$ such that

$$g_1\left(r \sum_{i=0}^{m} c_i\sigma^i\right) = r \sum_{i=0}^{m} c_i\sigma^i$$

or

$$g\left(r \sum_{i=0}^{m} c_i\sigma^i\right) = y(x^e) + \theta\sigma^*.$$

This implies that $x(t^e; (h!\,r)^{1/h}; c)$ is the endpoint of an admissible trajectory satisfying the boundary conditions with a smaller y_0 value, hence $u^0(t)$ is not optimal.

Actually $x(t^e)$ is not even a local minimum for we can choose θ as close to 0 as we choose subject to (3.5). Q.E.D.

4. Linear conditions for scalar controls. Suppose the control of (2.1) is a scalar and the set Ω is a subinterval of $\mathbb{R}$. The PMP characterizes the optimal control as one where the Hamiltonian achieves its maximum, and therefore we need only consider the endpoints of Ω and any interior points where (2.13) and (2.14) are satisfied. Typically for each x and λ this means considering only a finite number of discrete values of u. However, there is at least one important exception, namely, if the dynamics are linear in the control

$$\dot{x} = a_0(x) + ua_1(x). \tag{4.1}$$

Systems like this frequently arise in diverse applications because the assumption of linearity is so convenient in the formulation of mathematical models. Moreover, in Example 4.2, we show how necessary conditions developed for (4.1) can be easily extended to systems where the control enters nonlinearly.

If the dynamics is linear in u, then so is the Hamiltonian, H, and $\partial H/\partial u$ does not explicitly depend on u. If it is not zero, then the extremal control is bang-bang, i.e., at an endpoint of Ω. However, if it is zero, then the extremal control is singular since (2.14) is trivially satisfied. Moreover, (2.13) and (2.14) do not isolate the extremal control, and so we must consider the behavior of the system over an interval of time.

Suppose $u^0(t)$ and $x(t)$ are extremal for $t \in [t^0, t^e]$ for some choice of $\lambda(t)$. Assume that for $t \in (t^1, t^2)$, $u^0(t)$ is C^∞ and in the interior of Ω, hence singular. The Hamiltonian is given by

$$H(\lambda, x, u) = \lambda a_0(x) + u\lambda a_1(x),$$

and (2.13) reduces to

$$\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) = \lambda(t) a_1(x(t)) = 0 \tag{4.2}$$

for $t \in [t^1, t^2]$. Since $\lambda(t)$ annihilates H, this implies that

$$H(\lambda(t), x(t), u^0(t)) = \lambda(t) a_0(x(t)) = 0 \tag{4.3}$$

for $t \in [t^1, t^2]$.

It is straightforward to verify that given an arbitrary vector field $b(x)$ and any solution $\lambda(t)$ of the adjoint differential equation along the trajectory $x(t)$ which is generated by the control $u^0(t)$,

$$\frac{d}{dt}\lambda(t) b(x(t)) = \lambda(t)[a_0, b](x(t)) + u^0(t)\lambda(t)[a_1, b](x(t)), \tag{4.4}$$

where the Lie bracket is defined by

$$[a_i, b](x) = \left(\frac{\partial}{\partial x} b(x)\right) a_i(x) - \left(\frac{\partial}{\partial x} a_i(x)\right) b(x).$$

Repeated differentiation of (4.2) yields

$$\frac{d^k}{dt^k}\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) = 0 \tag{4.5}$$

for $t \in [t^1, t^2]$ and $k = 0, \cdots, \infty$. In particular,

$$\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) = \lambda(t) a_1(x(t)) = 0, \tag{4.6}$$

$$\frac{d}{dt}\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) = \lambda(t)[a_0, a_1](x(t)) = 0, \tag{4.7}$$

$$\begin{aligned} &\frac{d^2}{dt^2}\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) \\ &\quad = \lambda(t)[a_0[a_0, a_1]](x(t)) + u^0(t)\lambda(t)[a_1[a_0, a_1]](x(t)) = 0, \end{aligned} \tag{4.8}$$

$$\begin{aligned} \frac{d^3}{dt^3}\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) &= \lambda(t)[a_0[a_0[a_0, a_1]]](x(t)) \\ &+ 2u^0(t)\lambda(t)[a_0[a_1[a_0, a_1]]](x(t)) + (u^0(t))^2\lambda(t)[a_1[a_1[a_0, a_1]]](x(t)) \\ &+ \dot u^0(t)\lambda(t)[a_1[a_0, a_1]](x(t)) = 0, \end{aligned} \tag{4.9}$$

and so on. (In the next section we show that $[a_0[a_1[a_0, a_1]]] = [a_1[a_0[a_0, a_1]]]$.) One could also differentiate (4.3), however, no new conditions result.

Since $u^0(t) \in$ interior Ω for $t \in (t^1, t^2)$ we can, without loss of generality, assume that $u^0(t) = 0$ for $t \in [t^1, t^2]$ and $\pm 1 \in \Omega$ by redefining a_0 and a_1 as $a_0 + u^0 a_1$ and ca_1 for some constant c and by choosing a slightly smaller interval $[t^1, t^2]$ and a new Ω so that every admissible trajectory of the new system is also admissible for the old. Then (4.5) simplifies to

$$\frac{d^k}{dt^k}\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) = \lambda(t) ad^k(a_0)a_1(x(t)) = 0 \tag{4.10}$$

for $t \in [t^1, t^2]$, where $ad^0(a_0)a_1 = a_1$ and $ad^k(a_0)a_1 = [a_0, ad^{k-1}(a_0)a_1]$.

Equation (4.5) (or (4.10)) is sometimes referred to as the *linear necessary condition* for an optimal control because it is precisely this condition that one would obtain by linearizing (4.1) around the reference trajectory and considering the effect of a sequence of first order control variations at properly chosen times. This is the McShane–Pontryagin approach. Moreover in the case of (4.10), it involves brackets of a_0 and a_1 which are linear in the controllable vector field a_1.

Equation (4.3) is a constant necessary condition, i.e., it is zero order with respect to the controllable vector field a_1 of (4.1). It follows from the first order control variations

$$\alpha^{\pm}(s)x = \gamma^0(\pm s)x \tag{4.11}$$

whose derivative applied to (2.9) yields

$$\lambda(t)\frac{d}{ds}\alpha^{\pm}(0)x(t) = \pm\lambda(t)(a_0(x(t)) + u^0(t)a_1(x(t))) \leqq 0.$$

Since $x_0 = t$, this condition is independent of (4.5) (or 4.10)). Therefore $u^0(t) = 0$ is extremal for $t \in [t^1, t^2]$ if and only if the rank of $\{ad^k(a_0)a_1(x(t)) : k = 0, \cdots, \infty\}$ is less than n at each $t \in [t^1, t^2]$. (If the rank is n, then (4.3) and (4.10) supply $n+1$ linearly independent conditions, and hence only $\lambda(t) = 0$ satisfies them.)

Notice that (4.10) was not obtained directly from a control variation, but rather by differentiating (4.2). As a first application of the HMP, we would like to develop (4.10) directly via high order control variations. In the next section, these same control variations are used to obtain conditions which are quadratic in a_1.

Before we start, perhaps a word or two is required about terminology. When we speak of high order control variations, the order is with respect to the parameter s of the variations which is a time-like parameter. On the other hand, when we speak of linear or quadratic conditions, we mean relative to the controllable part of (4.1), i.e., of first or second order with respect to the integral of the absolute variation in control. In particular, when $u^0(t)=0$, these conditions can be expressed using brackets which are linear or quadratic in a_1.

Suppose $\gamma^{\pm 1}$ and γ^0 are the flows of $u^{\pm 1}(t)=\pm 1$ and $u^0(t)=0$. Then define the control variations

$$\alpha^{\pm 0}(s)x=\gamma^{\pm 1}(s)\gamma^0(-s)x.$$

Computing the first derivative,

$$\frac{d}{ds}\alpha^{\pm 0}(0)x=\pm a_1(x)=\pm ad^0(a_0)a_1(x),$$

and so these control variations yield (4.10) for $k=0$.

Next define

$$\alpha^{\pm 1}(s)x=\gamma^{\pm 1}(s)\gamma^0(s)\gamma^{\mp 1}(s)\gamma^0(-3s)x,$$

which are variations of order two, since $(d/ds)\alpha^{\pm 1}(0)x=0$.

To compute the second derivative, it is convenient to use the chain rule technique and allow $\alpha^{\pm 1}$ to operate on an arbitrary C^∞-function φ.

$$\begin{aligned}\frac{d^2}{ds^2}\varphi(\alpha^{\pm 1}(0)x)&=\pm 4(a_0a_1(\varphi(x))-a_1a_0(\varphi(x)))\\&=\pm 4[a_0,a_1]\varphi(x)=\pm 4ad(a_0)a_1(\varphi(x)).\end{aligned}$$

When applied to the HMP this yields (4.10) with $k=1$.

We generalize the above for any integer $r\geqq 1$. Define

$$\alpha_r^{\pm 1}(s)x=\gamma^{\pm 1}(s^r)\gamma^0(s)\gamma^{\mp 1}(s^r)\gamma^0(-s-2s^r)x.$$

If s^r is replaced by s_1 and s by s_2, then the chain rule implies that at $s=0$,

$$\begin{aligned}&\frac{d^j}{ds^j}=\frac{\partial^j}{\partial s_2^j},\\ (4.12)\quad&\frac{d^{r+j}}{ds^{r+j}}=\frac{(r+j)!}{j!}\frac{\partial^j}{\partial s_2^j}\frac{\partial}{\partial s_1}+\frac{\partial^{r+j}}{\partial s_2^{r+j}},\\&\frac{d^{2r+j}}{ds^{2r+j}}=\frac{(2r+j)!}{j!}\frac{(2r+j)!}{j!2!}\frac{\partial^j}{\partial s_2^j}\frac{\partial^2}{\partial s_1^2}+\frac{(2r+j)!}{j!}\frac{\partial^{r+j}}{\partial s_2^{r+j}}\frac{\partial}{\partial s_1}+\frac{\partial^{2r+j}}{\partial s_2^{2r+j}}\end{aligned}$$

for $j=0,\cdots,r-1$.

From this it follows that

$$\frac{d^j}{ds^j}\varphi(\alpha_r^{\pm 1}(0)x)=0$$

for $j=1,\cdots,r$ and

$$\frac{d^{r+1}}{ds^{r+1}}\varphi(\alpha_r^{\pm 1}(0)x)=\pm(r+1)!ad(a_0)a_1(\varphi(x)).$$

Of course $\alpha_r^{\pm 1}(s)x$ does not lead to a new condition, but its generalizations $\alpha_r^{\pm k}(s)x$ do, where for k odd,

$$\tag{4.13a} \begin{aligned}\alpha_r^{\pm k}(s)x=\gamma^{\pm 1}\left(\binom{k}{0}s^r\right)\gamma^0(s)\gamma^{\mp 1}\left(\binom{k}{1}s^r\right)\gamma^0(s)\\ \cdots\gamma^0(s)\gamma^{\mp 1}\left(\binom{k}{k}s^r\right)\gamma^0(-ks-2^ks^r)x,\end{aligned}$$

and for k even,

$$\tag{4.13b} \begin{aligned}\alpha_r^{\pm k}(s)x=\gamma^{\pm 1}\left(\binom{k}{0}s^r\right)\gamma^0(s)\gamma^{\mp 1}\left(\binom{k}{1}s^r\right)\gamma^0(s)\\ \cdots\gamma^0(s)\gamma^{\pm 1}\left(\binom{k}{k}s^r\right)\gamma^0(-ks-2^ks^r)x.\end{aligned}$$

Using (4.12) it can be shown that

$$\frac{d^j}{ds^j}\varphi(\alpha_r^{\pm k}(0)x)=0$$

for $j=1,\cdots,r-1$ and

$$\begin{aligned}\frac{d^{r+j}}{ds^{r+j}}\varphi(\alpha_r^{\pm k}(0)x)&=\pm\frac{(r+j)!}{j!}\sum_{l=0}^{k}\sum_{i=0}^{j}\binom{j}{i}l^i\binom{k}{l}(-l)^{j-i}a_0^{j-i}(a_0\pm(-1)^la_1)a_0^i(\varphi(x))\\ &=\pm\frac{(r+j)!}{j!}\sum_{i=0}^{j}\binom{j}{i}(-1)^{j-i}c_{j,k}a_0^{j-i}a_1a_0^i(\varphi(x)),\end{aligned}$$

where $c_{j,k}=\sum_{l=0}^{k}(-1)^l\binom{k}{l}l^j$.

In the next lemma we show that $c_{j,k}=0$ if $0\leqq j<k$ and $c_{k,k}=(-1)^kk!$ From this it follows that if $k<r$, then $\alpha_r^{\pm k}$ is a control variation of order $r+k$ and

$$\frac{d^{r+k}}{ds^{r+k}}\varphi(\alpha_r^{\pm k}(0)x)=\pm(-1)^k(r+k)!\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}a_0^{k-i}a_1a_0^i(\varphi(x)),$$

which by induction on k can be shown to equal

$$\pm(-1)^k(r+k)!ad^k(a_0)a_1(\varphi(x)).$$

Applying these variations to the HMP yields all the linear necessary conditions (4.10).

LEMMA 4.1.[1] *For any integers* $0 \leqq j \leqq k$, *let*

$$c_{j,k} = \sum_{l=0}^{k} (-1)^l \binom{k}{l} l^j.$$

Then $c_{j,k} = 0$ *if* $j < k$ *and* $c_{k,k} = (-1)^k k!$

Proof. By the binomial formula,

$$(e^t - 1)^k = \sum_{l=0}^{k} (-1)^{k-l} \binom{k}{l} e^{lt}.$$

Expanding e^{lt} in a Taylor series yields

$$(e^t - 1)^k = \sum_{j=0}^{\infty} \sum_{l=0}^{k} (-1)^{k-l} \binom{k}{l} l^j \frac{t^j}{j!}.$$

For $j = 0, \cdots, k-1$, the coefficient of t^j on the left is clearly 0 so

$$0 = \sum_{l=0}^{k} (-1)^{k-1} \binom{k}{l} \frac{l^j}{j!} = \frac{(-1)^k}{j!} c_{j,k}.$$

The coefficient of t^k is clearly 1 so

$$1 = \sum_{l=0}^{k} (-1)^{k-l} \binom{k}{l} \frac{l^k}{k!} = \frac{(-1)^k}{k!} c_{k,k}. \qquad \text{Q.E.D.}$$

Remark. In constructing $\alpha_r^{\pm k}(s)x$, we used the flows $\gamma^{\pm 1}$ of $a_0 \pm a_1$ to obtain a high order variation whose first nonzero derivative is a *multiple* of $\pm ad^k(a_0)a_1$ $(= ad^k(a_0)(a_0 \pm a_1))$. Suppose $\beta^{\pm}(s)x$ are control variations of order h along $x(t)$ whose h derivatives are $\pm b(x(t))$ for some vector field $b(x)$. Let $\beta^{\pm}(s)x = \gamma^{\pm j}(p_{\pm j}(s)) \cdots \gamma^{\pm 1}(p_{\pm 1}(s))\gamma^0(q_{\pm}(s))x$, where $\gamma^{\pm i}$ are flows of admissible controls and $p_{\pm i}(s) \geqq 0$ for small s. Define $\zeta^{\pm 1}(s)x = \gamma^{\pm j}(p_{\pm j}(s)) \cdots \gamma^{\pm 1}(p_{\pm 1}(s))x$, and construct $\alpha_r^{\pm k}(s)x$ as in (4.13) but with $\zeta^{\pm 1}$ replacing $\gamma^{\pm 1}$. If $k < r \cdot h$, the result is a control variation of order $k + r \cdot h$ whose $k + r \cdot h$ derivative is a multiple of $\pm ad^k(a_0)b(x(t))$ along $x(t)$.

Example 4.1. Consider the linear system

$$\dot{x} = A(t)x + ub(t),$$

where $x(0) = x0$ and $|u| \leqq 1$. Introduce time as a state variable, $x_0 = t$, so that the system is autonomous and define $\mathbf{x} = (x_0, x)$ such that

$$a_0(\mathbf{x}) = \begin{pmatrix} 1 \\ A(x_0)x \end{pmatrix}, \qquad a_1(\mathbf{x}) = \begin{pmatrix} 0 \\ b(x_0) \end{pmatrix}.$$

Then

$$[a_0, a_1](\mathbf{x}) = \begin{vmatrix} 0 \\ \dfrac{d}{dx_0} b(x_0) - A(x_0)b(x_0) \end{vmatrix},$$

$$ad^2(a_0)a_1(\mathbf{x}) = \begin{vmatrix} 0 \\ \dfrac{d^2}{dx_0^2} b(x_0) - \left(\dfrac{d}{dx_0} A(x_0)\right) b(x_0) - 2A(x_0) \dfrac{d}{dx_0} b(x_0) + A^2(x_0)b(x_0) \end{vmatrix},$$

[1] The author is indebted to H. Hermes for the proof of Lemma 4.1.

and so on. For autonomous systems this simplifies to

$$ad^k(a_0)a_1(\mathbf{x}) = \begin{pmatrix} 0 \\ (-1)^k A^k b \end{pmatrix}.$$

Any bracket which is homogeneous of degree two or more in $a_1(\mathbf{x})$ is identically zero. Therefore $\partial H/\partial u$ and all its time derivatives are independent of u, and (4.5) reduces to (4.10) regardless of whether $u^0(t) = 0$ or not.

Suppose the system is *controllable*, i.e., at each x there exists a k such that $a_0(\mathbf{x}), a_1(\mathbf{x}), \cdots, ad^k(a_0)a_1(\mathbf{x})$ is of full rank, $n+1$. Then there exists no nontrivial $\boldsymbol{\lambda}(t)$ satisfying (4.3) and 4.10) and any extremal control must be bang-bang, $|u(t)| = 1$. A similar analysis is given by Hermes and La Salle in § 9 of [19].

Example 4.2. Consider a nonlinear system which is *not necessarily* linear in the control

$$\dot{x} = f(x, u),$$

where $x(0) = x^0$ and $u \in \Omega$. Given a reference control $u^0(t) \in$ interior Ω for $t \in (t^1, t^2)$, we can put the system in the form (4.1) by *prolonging* the control. Define a new state $x_{n+1} = u - u^0(t)$ and a new control $v = \dot{x}_{n+1}$. Let $\mathbf{x} = (x, x_{n+1})$ and

$$a_0(\mathbf{x}) = \begin{pmatrix} f(x, u^0(x_0) + x_{n+1}) \\ 0 \end{pmatrix}, \qquad a_1(\mathbf{x}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

On the hypersurface $x_{n+1} = 0$, which includes the reference trajectory of the original problem

$$ad^k(a_0)a_1(\mathbf{x}) = \begin{pmatrix} -ad^{k-1}(f_0)f_u(x) \\ 0 \end{pmatrix}, \tag{4.14}$$

where $f_0(x) = f(x, u^0(x_0))$ and $f_u(x) = (\partial/\partial u)f(x, u^0(x_0))$. Notice that prolongation introduces a new linear direction $a_1(\mathbf{x}(t))$ and shifts the other linear directions by one $-a_0$ factor (4.14). In particular, if the original problem is linear in the control, $\dot{x} = f_0(x) + uf_u(x)$, and $u^0(t) = 0$, then prolongation essentially shifts $ad^{k-1}(f_0)f_u$ to $-ad^k(a_0)a_1$.

Consider the necessary conditions (4.3) and (4.5) for the prolonged problem where $\boldsymbol{\lambda} = (\lambda, \lambda_{n+1})$ and $\mathbf{H}(\boldsymbol{\lambda}, \mathbf{x}, v) = \boldsymbol{\lambda}(a_0(\mathbf{x}) + va_1(\mathbf{x}))$. The reference control is $v^0(t) = 0$ and for $k = 1$,

$$\frac{\partial}{\partial v}\mathbf{H}(\boldsymbol{\lambda}(t), \mathbf{x}(t), v^0(t)) = \boldsymbol{\lambda}(t)a_1(\mathbf{x}(t)) = 0,$$

which implies that $\lambda_{n+1}(t) = 0$, i.e., the prolonged adjoint variable lives on the original state space.

For $k > 1$,

$$\begin{aligned} 0 = \frac{d^k}{dt^k}\frac{\partial}{\partial v}\mathbf{H}(\boldsymbol{\lambda}(t), \mathbf{x}(t), v^0(t)) &= \boldsymbol{\lambda}(t)ad^k(a_0)a_1(\mathbf{x}(t)) \\ &= -\lambda(t)ad^{k-1}(f_0)f_u(x(t)) \\ &= -\frac{d^{k-1}}{dt^{k-1}}\frac{\partial}{\partial u}H(\lambda(t), x(t), u^0(t)), \end{aligned} \tag{4.15}$$

and so prolongation also shifts, by one time derivative, the linear necessary conditions of the original problem. Moreover (4.3) for the prolonged problem reduces to (4.3) for the original:

$$\begin{aligned} 0 = \mathbf{H}(\boldsymbol{\lambda}(t), \mathbf{x}(t), v^0(t)) &= \boldsymbol{\lambda}(t) a_0(\mathbf{x}(t)) \\ &= \lambda(t) f(x(t), u^0(t)) \\ &= H(\lambda(t), x(t), u^0(t)). \end{aligned}$$

Prolongation increases the dimension of the state space by one, but also introduces linear controllability in that direction $(a_1(x(t)))$ along the reference trajectory, so the codimension of linear controllability remains constant, and extremal trajectories remain extremal. Prolongation can also be viewed as restricting the class of admissible control variations. In the original problem the variation in u was required to be piecewise C^∞; in the prolonged problem the variation in u is continuous and piecewise C^∞. It is interesting to note that this smaller class yields the same necessary conditions for $u^0(t) \in$ interior Ω. This is a consequence of the infinite differentiability of the original problem and might not hold if it were only finitely differentiable.

Perhaps a word or two about the form of the control variations $\alpha_r^{\pm k}(s)x$ is in order. They somewhat resemble the variations of Kelley, Kopp and Moyer [8]. The derivatives of $\alpha_r^{\pm k}(s)x$ can be conveniently thought of as polynomials in the noncommuting variables a_0 and a_1. Parametrizing $\gamma^{\pm 1}$ by s^r and γ^0 by s has the following net effect; one must differentiate r times to obtain an a_1 factor, but only once for an a_0 factor. This allows us to control the relative degrees of a_0 and a_1.

The binomial coefficients and signs of $\gamma^{\pm 1}$ give $\alpha_r^{\pm k}(s)x$ the appearance of a kth order difference operator where γ^0 is the shift operator and $\gamma^{\pm 1}$ are the positive and negative evaluation operators. Needless to say, this is no coincidence since $ad^k(a_0)a_1(x)$ is precisely the kth time derivative of $a_1(x(t))$ in any coordinate system where $a_0(x)$ is a constant vector field. There are numerous other kth order difference operators and if one were to use them in an analogous fashion as models for constructing high order control variations, then the same necessary conditions would result.

It is not surprising that these necessary conditions can be expressed in terms of brackets of a_0 and a_1 for, as is well known [10], these brackets span all the directions in which the system (4.1) can evolve. However, it is a bit surprising that the $k+r$ derivative of $\alpha_r^{\pm k}$ should be exactly equal to a bracket of a_0 and a_1 when viewed as a formal polynomial in a_0 and a_1. There is a fundamental reason for this. Consider the real algebra of all formal polynomials in two noncommuting indeterminates, a_0 and a_1. The bracket is defined as before, $[a_0, a_1] = a_0a_1 - a_1a_0$. Then certain of these polynomials can be constructed from a_0 and a_1 via bracketing and forming linear combinations. Such polynomials are called *Lie elements* and they are characterized by *Friedrich's criterion* (see Jacobson [6, p. 170]) which, in our present context, can be described as follows. A formal polynomial in a_0 and a_1 is a Lie element if and only if whenever a_0 and a_1 are replaced by arbitrary C^∞-vector fields, the result is a first order partial differential operator on smooth functions, i.e., it involves only the first partial derivatives of the functions.

Since the first $r+k-1$ derivatives of $\alpha_r^{\pm k}$ are zero, Lemma (3.2) states that the $r+k$ derivative is a first order operator. Moreover, this is independent of the choice of a_0 and a_1, and so this derivative must be a Lie element. Because the degree of homogeneity in a_0 and a_1 is determined by the parametrization, the $r+k$ derivative can only be a multiple of $ad^k(a_0)a_1$, the only bracket which is homogeneous in a_0 and a_1 of the appropriate degrees.

A first order partial differential operator is characterized by the Liebnitz rule for the first derivative of a product of functions. (Friedrich's criterion is merely an abstract form of this.) The following will prove useful in the next section.

LEMMA 4.2. *Suppose $\alpha(s)x$ is a control variation of order h at x^1. Then the first $2h-1$ derivatives of $\alpha(s)x$ are first order partial differential operators on smooth functions at x^1.*

Proof. By Lemma (3.2) the first through h derivatives are first order operators at x^1. As for the others, let $\varphi(x)$ and $\psi(x)$ be smooth functions around x^1. Then by the generalized Liebnitz rule for higher derivatives,

$$\frac{d^j}{ds^j}\varphi\cdot\psi(\alpha(0)x^1)=\sum_{i=0}^{j}\binom{j}{i}\frac{d^i}{ds^i}\varphi(\alpha(0)x^1)\frac{d^{j-i}}{ds^{j-i}}\psi(\alpha(0)x^1).$$

Since $\alpha(s)x$ is of order h at x^1, only two terms of the right side are possibly nonzero if $0<j<2h$, so this reduces to the Liebnitz rule for first derivatives,

$$\frac{d^j}{ds^j}\varphi\cdot\psi(\alpha(0)x^1)=\varphi(x^1)\frac{d^j}{ds^j}\psi(\alpha(0)x^1)+\frac{d^j}{ds^j}\varphi(\alpha(0)x^1)\psi(x^1).\quad \text{Q.E.D.}$$

COROLLARY 4.3. *Suppose $\alpha(s)x$ is a control variation which is of order h at x^1 independent of the choice of a_0 and a_1. Then the first $2h-1$ derivatives must be Lie elements when viewed as formal polynomials in the indeterminates a_0 and a_1.*

Although the control variations considered in this section have not led to new necessary conditions, they are useful because their higher derivatives do, as we shall see in the next section. Another important aspect of these variations is that they allow us to make instantaneous control modifications to move in any linear direction. This property will allow us to cancel out undesirable lower order effects of other variations via Lemma 3.4, and thus arrive at higher order variations. We formalize this property in the following.

LEMMA 4.4. *Suppose $c_0(t),\cdots,c_k(t)$ are bounded C^∞-real-valued functions for $t\in(t^1,t^2)$. Define a vector field along $x(t)$ by*

$$b(t)=\sum_{i=0}^{k}c_i ad^i(a_0)a_1(x(t)).$$

Then for any $h>2k$, there exists a control variation $\beta(s)x$ of order h such that

$$\frac{d^h}{ds^h}\beta(0)x(t)=b(t)$$

for $t\in(t^1,t^2)$.

Proof. Proceed by induction on k. If $k=1$, *choose a constant* c large enough so that $|c_0(t)|\leq c$ for $t\in(t^1,t^2)$. Let $u^1(t)=c_0(t)/c$ and construct the control variations $\alpha_h^{+0}(s)x$ as before using $a_0+u^1a_1$ instead of a_0+a_1. This is a control variation of order h and

$$\frac{d^h}{ds^h}\alpha_h^{+0}(0)(x(t))=h!ad^0(a_0)(u^1a_1)(x(t))$$
$$=h!u^1(t)a_1(x(t))$$
$$=\frac{h!}{c}c_0(t)a_1(x(t)).$$

The desired variation is $\beta(s)x=\alpha_h^{+0}((c/h!)^{1/h}s)x$.

Now suppose the lemma is true for $k-1$. Then define $u^k(t)=c_k(t)/c$ where $c\geqq|c_k(t)|$ for $t\in(t^1,t^2)$. Construct $\alpha_r^{+k}(s)x$ using $a_0\pm(-1)^ku^ka_1$ instead of $a_0\pm a_1$ where $r=h-k>k$. This is a variation of order h and

$$\frac{d^h}{ds^h}\alpha_r^{+k}(0)(x(t))=h!ad^k(a_0)(u^ka_1)(x(t))$$
$$=h!u^k(t)ad^k(a_0)a_1(x(t))+\text{linear combination}$$
$$\text{of } ad^i(a_0)a_1(x(t))\quad\text{for } i=0,\cdots,k-1.$$

Define $\beta^k(s)x=\alpha_r^k((c/h!)^{1/h}s)x$. By induction there exist $\beta^{k-1}(s)x$ of order h such that

$$\frac{d^h}{ds^h}\beta^{k-1}(0)x(t)=b(t)-\frac{d^h}{ds^h}\beta^k(0)x(t).$$

The desired variation is obtained by "adding" β^k and β^{k-1} as described in Lemma 3.4. Q.E.D.

Remark. It is important to note that in the construction of these variations, the flow of $a_0\pm u^ia_1$ is parametrized by a multiple of s^r where $r=h-i$ for $i=0,\cdots,k$. Therefore should we continue to differentiate, the first derivative that could possibly involve a term quadratic in a_1 is the $2(h-k)$ derivative.

COROLLARY 4.5. *Consider the nonlinear system of Example* 4.2 *which is not necessarily linear in the control. Suppose $u^0(t)$ is C^∞ and in the interior of Ω for $t\in(t^1,t^2)$. Given any bounded C^∞-real-valued functions $c_0(t),\cdots,c_k(t)$ define a vector field along $x(t)$ by*

$$b(t)=\sum_{i=0}^{k}c_iad^i(f_0)f_u(x(t)).$$

Then for any $h>2k+2$, there exists a control variation $\beta(s)x$ of order h such that

$$\frac{d^h}{ds^h}\beta(0)x(t)=b(t)$$

for $t\in(t^1,t^2)$.

Proof. Prolong the problem as before and apply Lemma 4.4.

5. Quadratic conditions for scalar controls. For (4.1) assume that $u^0(t)=0\in$ interior Ω is an extremal control for $|t-t^1|<\varepsilon$. Then (2.14) is trivially satisfied so that $u^0(t)$ is singular. Moreover since $u^0(t)$ is an extremal control, (4.3) and (4.10) are satisfied for some $\lambda(t)$. Because these are equality constraints rather than inequality constraints, replacing $\lambda(t)$ with $-\lambda(t)$ does not alter them. Therefore they do not distinguish between minimizing and maximizing singular extremals.

To clear up this ambiguity, quadratic necessary conditions were developed by Kelley [7], Kopp and Moyer [9], Kelley, Kopp and Moyer [8], Tait [17], Goh [3], [4], Robbins [14], [15] and others. We refer the reader to the survey articles of Gabasov and Kirillova [2], Bell [1] and Jacobson [18] for extensive bibliographies. These conditions are sometimes referred to as the generalized Legendre–Clebsch conditions (GLC) because they resemble the Legendre–Clebsch condition (2.14) when expressed in terms of the Hamiltonian. Generally the proofs of the GLC ignore the problem of terminal constraints either by assuming there are not any, or by a normality assumption, a sometimes vague concept in the literature. Essentially, normality means that there exists sufficient local controllability around the reference trajectory to meet any terminal constraints that might be imposed without affecting the validity of the GLC. We give a more precise definition later.

In this section, using the HMP, we develop quadratic necessary conditions which generalize the GLC to problems with terminal constraints without using a normality assumption.

Let D_i^j denote the linear space of Lie elements which are homogeneous of degree i and j in the indeterminates a_0 and a_1, respectively, and let

$$D_i = \text{span} \bigcup_{j=0}^{\infty} D_i^j,$$

$$D^j = \text{span} \bigcup_{i=0}^{\infty} D_i^j,$$

$$D = \text{span} \bigcup_{i,j=0}^{\infty} D_i^j.$$

Let $D_i^j(x)(D_i(x), D^j(x), D(x))$ denote the linear subspace of a tangent vector at x obtained by substituting the vector fields of (4.1) in the Lie elements and evaluating at x.

Suppose $u^0(t)=0$ and $x(t)$ are a singular extremal control and trajectory on $[t^1, t^2]$. Following Robbins [15], we say that the control is *singular of degree* $h+1$ on this interval if h is the smallest integer such that for some $t \in (t^1, t^2)$,

$$[a_1, ad^h(a_0)a_1](x(t)) \notin D^1(x(t)).$$

The next theorem describes the quadratic necessary conditions for such a control to be minimal.

THEOREM 5.1. *Assume that $u^0(t)$ and $x(t)$ are defined for* (4.1) *on* $[t^0, t^e]$. *Suppose $u^0(t)=0 \in$ interior Ω on the subinterval (t^1, t^2). If u is singular of degree $h+1$ on this subinterval and h is finite, then h is odd. If $u^0(t)$ is minimal, then there exists a $\lambda(t)$ satisfying the PMP on $[t^0, t^e]$ such that*

$$(-1)^{(h+1)/2}\lambda(t)[a_1, ad^h(a_0)a_1](x(t)) \leqq 0$$

on the subinterval $[t^1, t^2]$.

Note that the theorem does not imply that the degree of singularity is finite, just that if $h<\infty$, then h must be odd, whether the extremal trajectory is minimal or not. Later we give an example where the degree of singularity is infinite. There may exist several subintervals of $[t^0, t^e)$ on which the degree of singularity varies. Before proving the theorem, we state a generalization and a corollary which do not assume linearity in the control or $u^0(t)=0$. First we make a generalized definition. Suppose $u^0(t)$ and $x(t)$ are a singular extremal control and trajectory for (2.1) on $[t^1, t^2]$. The control is *singular of degree* $h+1$ on this interval if h is the smallest integer for which there exists $\lambda(t)$ satisfying the adjoint differential equation (2.6) and the constant and linear necessary conditions

$$H(\lambda(t), x(t), u^0(t))=0, \tag{5.1}$$

$$\frac{d^k}{dt^k}\frac{\partial}{\partial u}H(\lambda(t), x(t), u^0(t))=0 \tag{5.2}$$

for $k=0, \cdots, \infty$ on any nontrivial subinterval of $[t^1, t^2]$ such that for some t in this subinterval,

$$\frac{\partial}{\partial u}\frac{d^{h+1}}{dt^{h+1}}\frac{\partial}{\partial u}H(\lambda(t), x(t), u^0(t))\neq 0.$$

Notice that a control could be singular of degree $h+1=0$ or ∞.

THEOREM 5.2. *Assume that $u^0(t)$ and $x(t)$ are defined for* (2.1) *on* $[t^0, t^e]$. *Suppose $u^0(t)\in$ interior Ω on the subinterval (t^1, t^2). If u is singular of degree $h+1$ on this subinterval and h is finite, then h is odd. If $u^0(t)$ is minimal, then there exists a $\lambda(t)$ satisfying the PMP on $[t^0, t^e]$ such that*

$$(-1)^{(h+1)/2}\frac{\partial}{\partial u}\frac{d^{h+1}}{dt^{h+1}}\frac{\partial}{\partial u}H(\lambda(t), x(t), u^0(t))\leqq 0$$

on the subinterval $[t^1, t^2]$.

A singular extremal control $u^0(t)\in$ interior Ω and trajectory $x(t)$ *are normal* on (t^1, t^2) if for each $t\in(t^1, t^2)$ there exists only one linearly independent $\lambda(t)$ satisfying the constant and linear necessary conditions (5.1) and (5.2). Since $x=(x_0, \cdots, x_n)$ this is equivalent to the assumption that the variations $\alpha^{\pm}(s)x$ of (4.11) and $\alpha_r^{\pm k}(s)x$ of (4.13) supply exactly n-dimensional local controllability at each $x(t)$ for $t\in(t^1, t^2)$.

In particular for (4.1) and $u^0(t)=0$, this is equivalent to the assumption that the dimension of $D^1(x(t))$ is $n-1$ for each $t\in(t^1, t^2)$.

COROLLARY 5.3 (Kelley, Kopp and Moyer [8]). *Assume that $u^0(t)$ and $x(t)$ are defined for* (2.1) *on* $[t^0, t^e]$. *Suppose $u^0(t)\in$ interior Ω and is normal on the subinterval $[t^1, t^2]$. If $u^0(t)$ is minimal, then there exists a $\lambda(t)$ satisfying the PMP on $[t^0, t^e]$ which is unique to the scalar multiple by normality. Let h be the smallest integer such that*

$$\frac{\partial}{\partial u}\frac{d^{h+1}}{dt^{h+1}}\frac{\partial}{\partial u}H(\lambda(t), x(t), u^0(t))\neq 0$$

for some $t \in (t^1, t^2)$. *If* h *is finite, then* h *is odd and on the subinterval* $[t^1, t^2]$,

$$(-1)^{(h+1)/2} \frac{\partial}{\partial u} \frac{d^{h+1}}{dt^{h+1}} \frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) \leqq 0$$

must hold.

Proof. We now prove Theorem 5.1. In Example (5.2) we show that Theorem 5.2 is equivalent to Theorem 5.1 by prolongation. The corollary follows immediately from Theorem 5.2.

Except for a nowhere dense set, every $t \in (t^1, t^2)$ is contained in an open interval where $D^1(x(t))$ is of constant dimension with a basis consisting of

$$\{a_1(x(t)), ad(a_0)a_1(x(t)), \cdots, ad^l(a_0)a_1(x(t))\}$$

for some l. Without loss of generality we can assume that the open interval is all of (t^1, t^2), for at other points the theorem follows by taking continuous limits.

By repeated application of the Jacobi identity for Lie elements, $[b_i[b_j, b_k]] = [[b_i, b_j]b_k] + [b_j[b_i, b_k]]$, and the skew symmetry relation, $[b_i, b_j] = -[b_j, b_i]$, it is easy to see that

$$\begin{aligned}[a_1, ad^i(a_0)a_1] = &\sum_{j=0}^{l-1} (-1)^j [a_0[ad^j(a_0)a_1, ad^{i-j-1}(a_0)a_1]] \\ &+ (-1)^l [ad^l(a_0)a_1, ad^{i-l}(a_0)a_1].\end{aligned} \tag{5.3}$$

If i is even and $l = i/2$, then skew symmetry implies that the last term on the right side is zero so

$$[a_1, ad^i(a_0)a_1] = \sum_{j=0}^{(i/2)-1} (-1)^j [a_0[ad^j(a_0)a_1, ad^{i-j-1}(a_0)a_1]]. \tag{5.4}$$

From (5.3) and (5.4) it can be shown that a basis for the linear space D_i^2 of Lie elements consists of

$$\{[a_1, ad^i(a_0)a_1], ad^2(a_0)[a_1, ad^{i-2}(a_0)a_1], \cdots, ad^{i-1}(a_0)[a_1[a_0, a_1]]\} \tag{5.5a}$$

if i is odd, and

$$\{[a_0[a_1, ad^{i-1}(a_0)a_1]], ad^3(a_0)[a_1, ad^{i-3}(a_0)a_1], \cdots, ad^{i-1}(a_0)[a_1[a_0, a_1]]\} \tag{5.5b}$$

if i is even.

Now suppose that u is singular of degree $h+1$, i.e.,

$$[a_1, ad^i(a_0)a_1](x(t)) \in D^1(x(t))$$

on the subinterval $[t^1, t^2]$ for $i = 1, \cdots, h-1$, but not for $i = h$. Bracketing both sides with a_0 yields

$$ad^j(a_0)[a_1, ad^i(a_0)a_1](x(t)) \in D^1(x(t)) \tag{5.6}$$

for $j = 0, \cdots, \infty$, $i = 1, \cdots, h-1$ and $t \in [t^1, t^2]$. In particular,

$$D_i^2(x(t)) \in D^1(x(t)) \tag{5.7}$$

on $[t^1, t^2]$ for $i = 1, \cdots, h-1$.

If h is even, (5.5b) and (5.6) imply that $[a_1, ad^h(a_0)a_1](x(t)) \in D^1(x(t))$ which contradicts the definition of h, hence h must be odd. Notice also that (5.6) implies that the only bracket which keeps (5.7) from being true for $i=h$ is $[a_1, ad^h(a_0)a_1](x(t))$.

To prove the rest of the theorem we must construct an appropriate high order control variation. We start with $\alpha_r^{+k}(s)x$ for any $k \geqq (h+1)/2$, $r>k$ and $r>h$. We have already computed the first $k+r$ derivatives of this variation and we know by Corollary 4.3 that the first $2(k+r)-1$ derivatives are Lie elements. We wish to study the jth derivative where $k+r<j\leqq h+2r\leqq 2(k+r)-1$.

It is easy to see that this variation causes no displacement in the time direction and so the jth derivative cannot possibly contain an a_0 term. Moreover, from the parametrization of the components of $\alpha_r^{+k}(s)x$ we know that the jth derivative is a sum of elements of D^1 if $r\leqq j<2r$ and a sum of elements of D^1 and D^2 if $2r\leqq j<3r$.

We already have control variations in the directions of $D^1(x(t))$ so we are only interested in the part of the jth derivative that lies in D^2 for $2r<j\leqq 3r$. Again from the parametrization we know that the part from D^2 is more precisely from D^2_{j-2r} and hence a linear combination of (5.5). Moreover from (5.7) we know that

$$D^2_{j-2r}(x(t)) \in D^1(x(t))$$

for $1\leqq j-2r<h$, so the first derivative that could possibly furnish a new test is $j=h+2r$. (Note that by the choice of r and k, it follows that $j<3r$ and $j<2(k+r)$ as desired.) This derivative can be expanded in the basis (5.5) but we are really only interested in the coefficient of $[a_1, ad^h(a_0)a_1]$ for this is the part of the derivative that lies outside $D^1(x(t))$.

To compute the coefficient of $[a_1, ad^h(a_0)a_1]$, we need only compute the coefficient of the monomial $a_1a_0^ha_1$ in the jth derivative for this is the only bracket of (5.5) that contains that monomial. We defer to a later lemma the computation that shows that the sign of this coefficient is $(-1)^{(h+1)/2}$.

In summary, we know the following. The first $k+r-1$ derivatives of $\alpha_r^{+k}(s)(x(t))$ are zero, derivatives $k+r$ through $h+2r-1$ lie in $D^1(x(t))$ and the $h+2r$ derivative consists of some parts from $D^1(x(t))$ plus a positive multiple of $(-1)^{(h+1)/2}[a_1, ad^h(a_0)a_1](x(t))$. To complete the proof we must make $\alpha_r^{+k}(s)x$ into a control variation of order $h+2r$ at $x(t)$ by canceling out all the lower derivatives for $t\in(t^1, t^2)$.

To do this we must apply Lemma (4.4) using the fact that $\{a_1(x(t)), \cdots, ad^l(a_0)a_1(x(t))\}$ spans $D^1(x(t))$. The lemma allows us to construct a control variation of any order $>2l$ whose first nonzero derivative is any vector field along $x(t)$ which lies in $D^1(x(t))$. Therefore we must choose k and r such that $k+r>2l$ and by "adding" new variations to $\alpha_r^{+k}(s)x$, we can cancel out its lower derivatives from $k+r$ through $h+2r-1$. Call the resulting variation $\beta(s)x$.

We must be careful in doing this, for it is possible that the sign of $[a_1, ad^h(a_0)a_1](x(t))$ in the $h+2r$ derivative of $\beta(s)x$ differs from its sign in the $h+2r$ derivative of $\alpha_r^{+k}(s)x$, and this would change the test. Recall that the parameters of the flows of $a_0\pm a_1$ in $\alpha_r^{+k}(s)x$ are s^r and, on the other hand, the variations of Lemma 4.4 used to cancel derivatives $r+k$ through $2r+h-1$ are composed of the flows of $a_0\pm u^ia_1$ parametrized by s^{r+k-l} or higher powers of s.

So, if $k>l$, then these "added" variations cannot possibly change the coefficient of $[a_1, ad^h a_0 a_1](x(t))$, although it could change the part of the $h+2r$ derivative which lies in $D^1(x(t))$. However this is not important for the test since (4.10) implies that $\lambda(t)$ annihilates $D^1(x(t))$.

In closing, we emphasize that this necessary condition is an inequality precisely because the bracket involved is quadratic in a_1. If we used $\alpha_r^{-k}(s)x$ instead of $\alpha_r^{+k}(s)x$ as a base for our high order variation, the same necessary condition would result because this is equivalent to replacing $a_1(x)$ by $-a_1(x)$ which leaves invariant the brackets quadratic in a_1. Using either of these variations, we have controllability in the direction $(-1)^{(h+1)/2}[a_1, ad^h(a_0)a_1](x(t))$, but not its negative. Q.E.D.

LEMMA 5.4. *Let $c_{k,h}$ be the coefficient of $a_1 a_0^h a_1(\varphi(x))$ in*

$$\frac{d^{2r+h}}{ds^{2r+h}}\varphi(\alpha_r^{+k}(0)x)$$

where $k \geqq (h+1)/2$, $r>k$ and $r>h$. Then

$$c_{k,h}=0 \quad \text{if } h=2,4,\cdots,2k-2$$

and

$$(-1)^{(h+1)/2}c_{k,h}>0 \quad \text{if } h=1,3,\cdots,2k-1.$$

Proof. By direct computation,

$$c_{k,h}=(2r+h)!\sum_{0\leqq i<j\leqq k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}\frac{(j-i)^h}{h!}.$$

If $h>0$ and is even, then

$$2c_{k,h}=(2r+h)!\sum_{i,j=0}^{k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}\frac{(j-i)^h}{h!}.$$

Expand $(e^{-t}-1)^k(e^t-1)^k$ by the binomial formula:

$$(e^{-t}-1)^k(e^t-1)^k=\sum_{i,j=0}^{k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}e^{(j-i)t}.$$

Expand $e^{(j-i)t}$ in a Taylor series:

$$(e^{-t}-1)^k(e^t-1)^k=\sum_{h=0}^{\infty}\sum_{i,j=0}^{k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}\frac{(j-i)^h}{h!}t^h.$$

For $h=2,4,\cdots,2k-2$, the coefficient of t^h on the left side is clearly 0 so

$$0=\sum_{i,j=0}^{k}(-1)^{i+j}\binom{k}{i}\binom{k}{j}\frac{(j-i)^h}{h!}=\frac{2}{(2r+h)!}c_{k,h},$$

and the first claim of the lemma has been shown.

If we assume h is a real variable, then for fixed k, $c_{k,h}$ is a sum of k exponentials. Therefore it has at most $k-1$ zeros, which we have just shown to be $h=2,4,\cdots,2k-2$. It follows that $c_{k,h}$ alternates signs at $h=1,3,\cdots,2k-1$

and, in particular, the sign of $c_{k,2k-1}$ must be the same as the coefficient of the largest exponential which is $(-1)^k=(-1)^{(h+1)/2}$. Q.E.D.

Example 5.1. Consider the problem of minimizing $x_4(t^e)$ subject to $x^0=0$, $x_0(t^e)=1$ and

$$\dot{x}_0=1, \qquad \dot{x}_3=x_1^2/2,$$
$$\dot{x}_1=u, \qquad \dot{x}_4=-x_2^2/2,$$
$$\dot{x}_2=x_1,$$

Clearly the trajectory determined by $u^0(t)=0$ is not optimal, but let us apply the previous theorems and corollary. This trajectory, $x(t)=(t,0,0,0,0)$ for $t\in[0,1]$, is a singular extremal since $\lambda(t)=(0,0,0,0,-1)$ satisfies the PMP (uniquely to scalar multiple) and $\partial^2/\partial u^2 H=0$.

A straightforward calculation shows that $[a_1, ad(a_0)a_1](x(t))\notin D^1(x(t))$, so the degree of singularity $h+1=2$ and we apply the test

$$-\lambda(t)[a_1, ad(a_0)a_1](x(t))\leqq 0$$

which is trivially satisfied. Therefore Theorem 5.1 does not rule out $u^0(t)=0$.

To understand the relationship between them, let us apply the other theorem and corollary. For some $t\in[0,1]$ (in fact, every t), the adjoint vector $\mu(t)=(0,0,0,1,0)$ satisfies the necessary conditions (5.1) and (5.2) and the adjoint differential equation on $[0,1]$. For any $t\in[0,1]$,

$$\frac{\partial}{\partial u}\frac{d^2}{dt^2}\frac{\partial}{\partial u}H(\mu(t), x(t), u^0(t))\neq 0,$$

so again $h+1=2$ and Theorem 5.2 only allows us to test if

$$-\frac{\partial}{\partial u}\frac{d^2}{dt^2}\frac{\partial}{\partial u}H(\lambda(t), x(t), u^0(t))\leqq 0,$$

which of course is trivially satisfied.

As for Corollary 5.3, it does not apply since the trajectory is not normal (dimension of $D^1(x(t))$ is $2<n-1=3$).

These quadratic necessary conditions failed to rule out an obviously nonminimal (in fact, maximal) trajectory because the problem was not given as a minimal realization (see Sussmann [16]). We are only interested in $x_0(t^e)$ and $x_4(t^e)$, so we define $y(x)=(y_0(x), y_1(x))=(x_4, x_0)$ as our output. It is clear that the x_3 coordinate is superfluous to the input-output description of the problem and may be dropped. Then $[a_1, ad(a_0)a_1](x(t))=0$ and so $h+1=4$. Since

$$\lambda(t)[a_1, ad^3(a_0)a_1](x(t))=1\not\leqq 0,$$

the trajectory is nonoptimal. Similarly, when applying Theorem 5.2, we find $h=3$ and the corresponding test rules out $u^0(t)=0$. Moreover since the dimension $n+1$ of the state space is now 4 rather than 5, and the dimension of $D^1(x(t))$ is still 2, the trajectory is normal by the comments following Theorem 5.2. Corollary 5.3 also rules out the trajectory.

Notice that if an additional terminal constraint, $x_3(t^e)=0$, is added to the original problem, then the output map must be expanded to include x_3 and this coordinate cannot be eliminated. Therefore the quadratic necessary conditions no longer rule out the control $u^0(t)=0$, but this is as it should be for this control generates the only trajectory satisfying the terminal constraints.

Example 5.2. Once again consider a nonlinear system which is not necessarily linear in the control as in Example 4.2. Prolong the system as before by introducing new state and control variables. On the hypersurface $x_{n+1}=0$, which includes the reference trajectory

(5.8)
$$[a_1, ad^h(a_0)a_1](\mathbf{x}) = \begin{pmatrix} -ad^{h-1}(f_0)f_{u,u}(x) - \sum_{i=0}^{h-2} ad^{h-2-i}(f_0)[f_u, ad^i(f_0)f_u](x) \\ 0 \end{pmatrix}$$

when f_0, f_u are as before and $f_{u,u}(x)=(\partial^2/\partial u^2)f(x, u^0(x_0))$. This shows that the GLC for $h=1$ of the prolonged problem is equivalent to the Legendre–Clebsch condition of the original.

If the original problem is linear in the control, $\dot{x}=f_0(x)+uf_u(x)$, and $u^0(t)=0$, then the prolongation shifts $D^2_{h-2}(x(t))$ to $D^2_h(\mathbf{x}(t))$. It also shifts the GLC for $h-2$ to the GLC for h; the lower order GLC is satisfied with equality and therefore multiplication by $\boldsymbol{\lambda}(t)$ cancels all but $\lambda(t)[f_u, ad^{h-2}(f_0)f_u](x(t))$ on the right side of (5.8).

Prolongation can be used to show that Theorem 5.1 implies Theorem 5.2 for an arbitrary nonlinear system (2.1). A straightforward calculation shows that for $h \geqq 1$:

$$\begin{aligned} &\frac{\partial}{\partial u}\frac{d^{h-1}}{dt^{h-1}}\frac{\partial}{\partial u} H(\lambda(t), x(t), u^0(t)) \\ &\quad = \lambda(t)ad^{h-1}(f_0)f_{u,u}(x) + \sum_{i=0}^{h-2} \lambda(t)ad^i(f_0)[f_u, ad^{h-2-i}(f_0)f_u](x(t)) \\ &\quad = -\boldsymbol{\lambda}(t)[a_1, ad^h(a_0)a_1](\mathbf{x}(t)). \end{aligned}$$

Many proofs of the GLC are based on the reverse of prolongation. In the literature this is known as a "transformation of control variable" or "passing to the accessory minimum problem". By using the integral of the old control as the new control variable, the GLC for h is converted into the GLC for $h-2$. Repeated application reduces the GLC for h to the Legendre–Clebsch condition which has been previously demonstrated. The principal difficulty in applying this technique is that in effect, one is dropping the dimension of the state by changing a state variable to a control variable. Another way of looking at this is to say that one is allowing impulse controls. One must justify the claim that necessary conditions developed using this wider class of controls are also necessary conditions for the original class. These problems are usually ignored in the literature and instead normality is assumed to be sufficient to overcome any difficulties that arise in this fashion and also to meet the terminal constraints.

Example 5.3. To see that the degrees of singularity $h+1$ can be infinite, consider the problem of minimizing $x_2(t^e)$ subject to $x(0)=0$, $x_0(t^e)=1$, $|u|\leqq 1$ and

$$\dot{x}_0=1, \qquad \dot{x}_2=x_1^3/6,$$
$$\dot{x}_1=u+x_1^2.$$

A straightforward computation shows that along the trajectory $x(t)=(t,0,0)$ generated by the control $u^0(t)=0$ for the adjoint variable $\lambda(t)=(0,0,-1)$,

$$\frac{d^h}{dt^h}\frac{\partial}{\partial u}H(\lambda(t),x(t),u^0(t))=0$$

and

$$\frac{\partial}{\partial u}\frac{d^h}{dt^h}\frac{\partial}{\partial u}H(\lambda(t),x(t),u^0(t))=0$$

for all h. Therefore this trajectory is a singular extremal, and the quadratic tests are inconclusive. To show the nonoptimality of this trajectory, the HMP must be used to construct a necessary condition which is particularly suited to the problem at hand, i.e., a cubic or higher test.

Let $\alpha^{\pm}(s)x$ and $\alpha_1^{\pm 0}(s)x$ be defined by (4.11) and (4.13) respectively, such that

$$\frac{d}{ds}\alpha^{\pm}(0)x=\pm(1,0,0),$$

$$\frac{d}{ds}\alpha_1^{\pm 0}(s)x=\pm(0,1,0),$$

and so the trajectory is normal. Consider $\alpha_1^{\pm 1}(s)x$. It is easy to show that this is a variation of order two and

$$\frac{d^2}{ds^2}\alpha_1^{\pm 1}(s)x=\pm 4[a_0,a_1](x),$$

but this bracket is zero along $x(t)$. Therefore, in this case, $\alpha_1^{\pm 1}$ are of order at least 3. Computing the next derivative which also must be a Lie element (since $\alpha_1^{\pm 1}$ is of order 2 for all a_0 and a_1),

$$\frac{d^3}{ds^3}\alpha_1^{\pm 1}(s)x=\pm 6[a_0[a_0,a_1]](x)-10[a_1[a_0,a_1]](x).$$

The first of these brackets is zero and the second is $(0,-2,0)$ along $x(t)$. Therefore this derivative can be canceled by "adding" $\alpha_1^{\mp 0}(2^{1/3}s^3/6)x$ via Lemma 3.4. Call the resulting variation $\beta^{\pm}(s)x$; it is of order 4 along $x(t)$. Because of the parametrization of $\alpha_1^{\mp 0}$, it follows that the fourth derivatives of $\beta^{\pm}$ and $\alpha_1^{\pm 1}$ are identical. This derivative is not a Lie element, but it is in the span of $D(x(t))$. By direct computation,

$$\frac{d^4}{ds^4}\beta^{\pm}(s)x=\pm(0,0,6).$$

Applying this to the HMP yields the nonoptimality of $u^0(t)=0$.

This is actually a third order test because it is a multiple of $[a_1[a_1[a_0,a_1]]]$ $x)=(0,0,1)$ which is cubic in a_1. A similar conclusion can be obtained by the method of Hermes [20].

This example demonstrates how one constructs necessary conditions which are adapted to a particular problem. First one applies the standard linear and quadratic tests developed in the last two sections. If these are inconclusive, i.e., they are satisfied with equality rather than strict inequality, then one must construct cubic and higher tests ad hoc. The basic building blocks are the family of variations $\alpha_r^{\pm k}(s)x$. One works with as small a k and r as possible and considers the higher derivatives. If these can be canceled using linear or quadratic controllability, then one "adds" the appropriate variations to do so. Higher derivatives are considered and hopefully a definitive test is eventually realized.

6. Quadratic conditions for vector controls. In this section, we generalize the results of the last to a system with vector controls,

$$\dot{x} = a_0(x) + \sum_{i=1}^{l} u_i a_i(x), \tag{6.1}$$

where $u = (u_1, \cdots, u_l)$ is constrained to lie in Ω, a subset of $\mathbb{R}^l$ with nonempty interior.

By fixing all but one of the controls and varying the other, one obtains the previously discussed linear and quadratic necessary conditions for each control.

These are the same linear necessary conditions involving $(d^h/dt^h)(\partial/\partial u_i)H$ (or if $u^0(t) = 0$, $ad^h(a_0)a_i$) for $i = 1, \cdots, l$, and the same quadratic conditions involving $(\partial/\partial u_i)(d^h/dt^h)(\partial/\partial u_i)H$ (or if $u^0(t) = 0$, $[a_i, ad^h(a_0)a_i]$) for $i = 1, \cdots, l$. There are, however, new quadratic necessary conditions associated with the mixed partials $(\partial/\partial u_i)(d^h/dt^h)(\partial/\partial u_j)H$ (or if $u^0(t) = 0$, $[a_i, ad^h(a_0)a_j]$) for $i, j = 1, \cdots, l$.

These conditions were first developed by Goh [4] using a sequence of accessory minimum problems under an assumption of normality. We use the HMP to prove these results and extend them to problems with terminal constraints without normality.

Let D_i^j denote the linear space of Lie elements which are homogeneous of degree i in the indeterminate a_0 and homogeneous of degree j in the vector of indeterminates $(a_1, \cdots, a_l)$, and let D_i, D^j, D, $D_i^j(x)$, $D_i(x)$, $D^j(x)$ and $D(x)$ be as before. Since there is more than one controllable vector field, there are some significant differences. For example, D_0^2 which previously was $\{0\}$ since $[a_1, a_1] = 0$, now contains the nontrivial Lie elements $[a_i, a_j]$ where $i \neq j$.

Suppose the reference control is $u^0(t) = 0$. Associated with each control u_i, there is a *degree of singularity* $h_i + 1$ defined as before; h_i is the smallest integer such that for some $t \in (t^1, t^2)$,

$$[a_i, ad^{h_i}(a_0)a_i](x(t)) \notin D^1(x(t)).$$

We wish to emphasize the fact that $D^1(x(t))$ contains $ad^k(a_0)a_j(x(t))$ where $j \neq i$, but the arguments of Theorem 5.1 are still valid, so that if $h_i < \infty$, it must be odd.

THEOREM 6.1. *Assume that $u^0(t)$ and $x(t)$ are defined for* (6.1) *on $[t^0, t^e]$. Suppose $u^0(t) = 0 \in$ interior Ω and each u_i is singular of degree $h_i + 1$ on the subinterval (t^1, t^2). If $u^0(t)$ is minimal, then there exists a $\lambda(t)$ satisfying the PMP on $[t^0, t^e]$ such that on the subinterval $[t^1, t^2]$,*

$$\lambda(t)[a_i, ad^k(a_0)a_j](x(t)) = 0 \tag{6.2}$$

for $k=0,\cdots,(h_i+h_j)/2-1$, $1\leqq i,j\leqq l$. *Moreover, if* $h_i<\infty$ *for* $i=1,\cdots,k\leqq l$, *then the* $k\times k$ *matrix whose* i,j *entry is*

$$(6.3)\qquad (-1)^{(h_j+1)/2}\lambda(t)[a_i, ad^{(h_i+h_j)/2}(a_0)a_j](x(t)),$$

where $1\leqq i,j\leqq k$ *must be symmetric and nonpositive definite.*

The following theorem generalizes the above to an arbitrary nonlinear system (2.1) where $u^0(t)$ is not necessarily zero. Recall that the control u_i is singular of degree h_i+1 on $[t^1,t^2]$ if h_i is the smallest integer such that for some $t\in(t^1,t^2)$ there exists $\lambda(t)$ satisying the adjoint differential equation (2.6) and the constant and linear necessary condition

$$(6.4)\qquad H(\lambda(t),x(t),u^0(t))=0,$$

$$(6.5)\qquad \frac{d^k}{dt^k}\frac{\partial}{\partial u_j}H(\lambda(t),x(t),u^0(t))=0$$

for $k=0,\cdots,\infty$ and $j=1,\cdots,l$ on any nontrivial subinterval of $[t^1,t^2]$ such that for some t in this subinterval,

$$\frac{\partial}{\partial u_i}\frac{d^{h+1}}{dt^{h+1}}\frac{\partial}{\partial u_i}H(\lambda(t),x(t),u^0(t))\neq 0.$$

Again we wish to emphasize that (6.5) must hold for every u_j and if $h_i<\infty$, it must be odd.

THEOREM 6.2. *Assume that* $u^0(t)$ *and* $x(t)$ *are defined for* (2.1) *on* $[t^0,t^e]$. *Suppose* $u^0(t)\in\text{interior }\Omega$ *and each* u_i *is singular of degree* h_i+1 *on the subinterval* (t^1,t^2). *If* $u^0(t)$ *is minimal, then there exists a* $\lambda(t)$ *satisfying the PMP on* $[t^0,t^e]$ *such that on the subinterval* $[t^1,t^2]$,

$$\frac{\partial}{\partial u_i}\frac{d^k}{dt^k}\frac{\partial}{\partial u_j}H(\lambda(t),x(t),u^0(t))=0$$

for $k=0,\cdots,(h_i+h_j)/2$, $1\leqq i,j\leqq l$. *Moreover if* $h_i<\infty$ *for* $i=1,\cdots,k\leqq l$, *then the* $k\times k$ *matrix whose* i,j *entry is*

$$(-1)^{(h_j+1)/2}\frac{\partial}{\partial u_i}\frac{d^{(h_i+h_j)/2+1}}{dt^{(h_i+h_j)/2+1}}\frac{\partial}{\partial u_j}H(\lambda(t),x(t),u^0(t)),$$

where $1\leqq i,j\leqq k$, *must be symmetric and nonpositive definite.*

As before, a singular extremal control $u^0(t)\in\text{interior }\Omega$ and trajectory $x(t)$ are *normal* on (t^1,t^2) if for each $t\in(t^1,t^2)$ there exists only one linearly independent vector $\lambda(t)$ satisfying the constant and linear necessary conditions (6.4) and (6.5).

COROLLARY 6.3 (Goh [4]). *Assume that* $u^0(t)$ *and* $x(t)$ *are defined for* (2.1) *on* $[t^0,t^e]$. *Suppose* $u^0(t)\in\text{interior }\Omega$ *and is normal on the subinterval* (t^1,t^2). *If* $u^0(t)$ *is minimal, then there exists a* $\lambda(t)$ *satisfying the PMP on* $[t^0,t^e]$, *which is unique to the scalar multiple by normality. Let* h_i *be the smallest integer such that*

$$\frac{\partial}{\partial u_i}\frac{d^{h_i+1}}{dt^{h_i+1}}\frac{\partial}{\partial u_i}H(\lambda(t),x(t),u^0(t))\neq 0$$

for some $t \in (t^1, t^2)$. *Then each finite* h_i *must be odd and on the subinterval* $[t^1, t^2]$ *such that*

$$\frac{\partial}{\partial u_i}\frac{d^k}{dt^k}\frac{\partial}{\partial u_j}H(\lambda(t), x(t), u^0(t)) = 0$$

must hold for $k = 0, \cdots, (h_i + h_j)/2$, $1 \leqq i, j \leqq l$. *Moreover if* $h_i < \infty$ *for* $i = 1, \cdots, k \leqq l$, *then the* $k \times k$ *matrix whose* i, j *entry is*

$$(-1)^{(h_j+1)/2}\frac{\partial}{\partial u_i}\frac{d^{(h_i+h_j)/2+1}}{dt^{(h_i+h_j)/2+1}}\frac{\partial}{\partial u_j}H(\lambda(t), x(t), u^0(t)),$$

where $1 \leqq i, j \leqq k$ *must be symmetric and nonpositive definite.*

Remark 1. Goh [4] does not express his necessary conditions in above form, but they are equivalent. The Hamiltonian formulation of Corollary 6.3 makes the conditions easier to describe and apply. They are closer to Robbins [15], but his results are weaker for they do not include quadratic conditions involving two controls which are singular of differing degrees.

Remark 2. The above results make it desirable to choose coordinates in the control space $\Omega \subseteq \mathbb{R}^l$ so that the controls $u_1, \cdots, u_l$ are singular of as high a degree as possible. We discuss how this is done in Examples 6.1 and 6.2.

Proof. We now prove Theorem 6.1, Theorem 6.2 follows by prolongation as before, and the corollary follows immediately from Theorem 6.2.

By repeated application of the Jacobi identity,

$$\begin{aligned}[a_i, ad^k(a_0)a_j] = \sum_{\sigma=0}^{\rho-1} (-1)^\sigma[a_0[ad^\sigma(a_0)a_i, ad^{k-\sigma-1}(a_0)a_j]] \\ + (-1)^\rho[ad^\rho(a_0)a_i, ad^{k-\rho}(a_0)a_j].\end{aligned} \tag{6.6}$$

Letting $\rho = k$, we obtain

$$\begin{aligned}[a_i, ad^k(a_0)a_j] = (-1)^{k+1}[a_j, ad^k(a_0)a_i] \\ + \sum_{\sigma=0}^{k-1} (-1)^\sigma[a_0[ad^\sigma(a_0)a_i, ad^{k-\sigma-1}(a_0)a_j]].\end{aligned} \tag{6.7}$$

These equations imply that a basis for the linear space D_k^2 of Lie elements consists of the union of (5.5) with

$$\{ad^{k-\sigma}(a_0)[a_i, ad^\sigma(a_0)a_j]: 0 \leqq \sigma \leqq k, 1 \leqq i < j \leqq k\}. \tag{6.8}$$

Note the presence of terms like $[a_i, ad^k(a_0)a_j]$ in this basis even when k is even.

If $i = j$, then (6.2) follows immediately because u_i is assumed singular of degree $h_i + 1$. Suppose $i \neq j$ and $k \leqq (h_i + h_j)/2 - 1$. Choose any k_i and k_j such that $k_i + k_j = k$ and $k_i \leqq (h_i - 1)/2$, $k_j \leqq (h_j - 1)/2$. Choose $r_i > k_i$, $r_j > k_j$ such that $k_i + r_i = k_j + r_j$. Define a pair of control variations

$$\beta^\pm(s)x = \zeta_{r_i}^{\pm k_i}(s)\xi_{r_j}^{+k_j}(s)\zeta_{r_i}^{\mp k_i}(s)\xi_{r_j}^{-k_j}(s)\gamma^0(-2p(s) - 2q(s))x,$$

where

$$\zeta_{r_i}^{\pm k_i}(s)x = \alpha_{r_i}^{\pm k_i}(s)\gamma^0(p(s))x$$

using the control a_i instead of a_1;

$$\xi_{r_j}^{\pm k_j}(s)x=\alpha_{r_j}^{\pm k_j}(s)\gamma^0(q(s))x$$

using the control a_j and $p(s)=k_is+2^{k_i}s^{r_i}$, $q(s)=k_js+2^{k_j}s^{r_j}$. In other words, $\beta^{\pm}(s)x$ are constructed by "adding" $\alpha_{r_i}^{+k_i}(s)x$ and $\alpha_{r_i}^{-k_i}(s)x$ made with a_i to $\alpha_{r_j}^{+k_j}(s)x$ and $\alpha_{r_j}^{-k_j}(s)x$ made with a_j.

From the definitions of $\zeta_{r_i}^{\pm k_i}$ and $\xi_{r_j}^{\pm k_j}(s)x$ and the chain rule we have

$$\frac{d^\rho}{ds^\rho}\zeta_{r_i}^{\pm k_i}(0)x=\frac{d^\rho}{ds^\rho}\gamma^0(p(0))x,$$

$$\frac{d^\rho}{ds^\rho}\xi_{r_j}^{\pm k_j}(0)x=\frac{d^\rho}{ds^\rho}\gamma^0(q(0))x$$

if $1\leqq\rho<k_i+r_i$, and

$$\frac{d^{k_i+r_i}}{ds^{k_i+r_i}}\zeta_{r_i}^{\pm k_i}(0)x=\frac{d^{k_i+r_i}}{ds^{k_i+r_i}}\gamma^0(p(0))x\pm(-1)^{k_i}\frac{(r_i+k_i)!}{k_i!}ad^{k_i}(a_0)a_i(x),$$

$$\frac{d^{k_i+r_i}}{ds^{k_i+r_i}}\xi_{r_j}^{\pm k_j}(0)x=\frac{d^{k_i+r_i}}{ds^{k_i+r_i}}\gamma^0(q(0))x\pm(-1)^{k_j}\frac{(r_j+k_j)!}{k_j!}ad^{k_j}(a_0)a_j(x).$$

Using this we see that

$$\begin{aligned}\frac{d^\rho}{ds^\rho}\beta^{\pm}(0)x&=\frac{d}{ds^\rho}\alpha_{r_i}^{\pm k_i}(0)x+\frac{d}{ds^\rho}\gamma^0(p(0))\alpha_{r_j}^{+k_j}(s)\gamma^0(-p(0))x\\&+\frac{d}{ds^\rho}\gamma^0(p(0)+q(0))\alpha_{r_i}^{\mp k_i}(0)\gamma^0(-p(0)-q(0))x\\&+\frac{d}{ds^\rho}\gamma^0(2p(0)+q(0))\alpha_{r_j}^{-k_j}(0)\gamma^0(-2p(0)-q(0))x\end{aligned}\tag{6.9}$$

if $1\leqq\rho<2(k_i+r_i)$. If $\rho=2(k_i+r_i)$ we have the above plus the extra terms on the right,

$$\mp(-1)^{k_i+k_j}\frac{((k_i+r_i)!)^2}{k_i!k_j!}[ad^{k_i}(a_0)a_i, ad^{k_j}(a_0)a_j].\tag{6.10}$$

We wish to make $\beta^{\pm}(s)x$ into a control variation of order $2(k_i+r_i)$ by "adding" other variations to cancel out the right side of (6.9) for $\rho=1,\cdots,2(k_i+r_i)$.

Since $\alpha_{r_i}^{\pm k_i}(s)x$ and $\alpha_{r_j}^{\pm k_j}(s)x$ are "added" to make $\beta^{\pm}(s)x$ and the former are of order k_i+r_i, so must be the latter. Moreover the k_i+r_i derivative of $\beta^{\pm}(s)x$ is just the sum of the corresponding derivatives of the former and hence zero. Therefore $\beta^{\pm}(s)x$ are control variations of order at least k_i+r_i+1 independent of a_0, a_i and a_j and so all their derivatives up to $2(k_i+r_i)+1$ are Lie elements.

Studying the right side of (6.9) for $\rho\leqq 2(k_i+r_i)$ we see that it contains no cross terms, i.e., terms with both an a_1 and an a_2 factor. It involves only linear brackets and brackets either quadratic in a_i or quadratic in a_j. By the relationship of k_i *and* k_j to the degrees of singularity h_i and h_j, it follows that the right side of (6.9) along $x(t)$ is in $D^1(x(t))$. Therefore it can be canceled using Lemma (4.4).

The result is a pair of control variations of order $2(k_i+r_i)$ whose $2(k_i+r_i)$ derivatives are given by (6.10). Using (2.9) of the HMP it follows that if $u^0(t)$ is minimal, then

$$\lambda(t)[ad^{k_i}(a_0)a_i, ad^{k_j}(a_0)a_j](x(t))=0. \tag{6.11}$$

Differentiating (6.11) with respect to time yields

$$\lambda(t)ad^p(a_0)[ad^{k_i}(a_0)a_i, ad^{k_j}(a_0)a_j](x(t))=0,$$

and so (6.6) and induction on $k=k_i+k_j$ implies (6.2).

To show (6.3) we first assume that each $h_i=h$ for $i=1,\cdots,k$. Let $M(t)$ be the $k\times k$ matrix whose i,j entry is

$$(-1)^{(h+1)/2}\lambda(t)[a_i, ad^h(a_0)a_j](x(t)).$$

It follows from (6.2) and (6.7) for odd h that $M(t)$ is symmetric. To show that $M(t)$ is nonpositive definite along minimal trajectories is equivalent to showing that for any $u(t)=(u_1(t),\cdots,u_k(t))$,

$$u(t)^TM(t)u(t)\leqq 0$$

along minimal trajectories.

Given any such $u(t)$, define a new system with scalar control v by

$$\dot{x}=b_0(x)+vb_1(x), \tag{6.12}$$

where $b_0(x)=a_0(x)$ and $b_1(x)=\sum u_i(x_0)a_i(x)$. Applying Theorem 5.1 to (6.12) we see that the control v is singular of degree $h+1$ and we obtain the necessary condition

$$(-1)^{(h+1)/2}\lambda(t)[b_1, ad^h(b_0)b_1](x(t))\leqq 0. \tag{6.13}$$

Now

$$\begin{aligned}[b_1, ad^h(b_0)b_1](x(t))&=\sum_{i,j} u_i(x_0)u_j(x_0)[a_i, ad^h(a_0)a_j](x(t))\\ &\quad+\text{terms from } D_\rho^2(x(t)) \quad \text{for } \rho<h.\end{aligned} \tag{6.14}$$

So, applying (6.13) and (6.2) to (6.14) yields the desired result.

Suppose the degrees of singularity of the various controls are not the same. If we prolong u_i, i.e., define a new state $x_{n+1}=u_i$ and new control $v=\dot{x}_{n+1}$, then we obtain a system of the form

$$\dot{\mathbf{x}}=b_0(\mathbf{x})+vb_i(\mathbf{x})+\sum_{j\neq i} u_jb_j(\mathbf{x}),$$

where $\mathbf{x}=(x, x_{n+1})$ and

$$b_0(\mathbf{x})=\begin{pmatrix}a_0(x)+x_{n+1}a_i(x)\\ 0\end{pmatrix},\qquad b_i(\mathbf{x})=\begin{pmatrix}0\\1\end{pmatrix},\qquad b_j(\mathbf{x})=\begin{pmatrix}a_j(x)\\0\end{pmatrix}.$$

Along $x_{n+1}=0$ for $j\neq i$,

$$[b_i, ad^\rho(b_0)b_i](\mathbf{x})=\begin{pmatrix}-\sum_{\sigma=0}^{\rho-2} ad^{\rho-\sigma-2}(a_0)[a_i, ad^\sigma(a_0)a_i](x)\\ 0\end{pmatrix},$$

$$[b_j, ad^\rho(b_0)b_j](\mathbf{x})=\begin{pmatrix}[a_j, ad^\rho(a_0)a_j(x)]\\ 0\end{pmatrix},$$

$$[b_i, ad^\rho(b_0)b_j](\mathbf{x})=\begin{pmatrix}\sum_{\sigma=0}^{\rho-1} ad^{\rho-\sigma-1}(a_0)[a_i, ad^\sigma(a_0)a_j](x)\\ 0\end{pmatrix},$$

$$[b_j, ad^\rho(b_0)b_i](\mathbf{x})=\begin{pmatrix}-[a_j, ad^{\rho-1}(a_0)a_i](x)\\ 0\end{pmatrix}.$$

Therefore the degree of singularity of v us h_i+3. In this way, all the controls can be made singular of the same degree and (6.3) follows from repeated use of the above identities. Q.E.D.

Example 6.1. Suppose $u^0(t)$ generates a normal singular extremal for (2.1) on $[t^1, t^2]$ and we wish to apply Corollary 6.3. To obtain as many necessary conditions as possible it is desirable to make a time-dependent change of coordinates of the control space $\Omega\subseteq\mathbb{R}^l$.

Start with the symmetric $l\times l$ matrix

$$M_0(t)=\left(\frac{\partial^2}{\partial u_i\partial u_j}H(\lambda(t), x(t), u^0(t))\right),$$

where $1\leqq i, j\leqq l$ and $\lambda(t)$ is uniquely determined to the scalar multiple by normality. By passing to a subinterval if necessary, we can assume that rank $M_0(t)$ is constant and equal to $l-l_1$ where $l_1\neq 0$ by assumption. There exists an orthonormal basis $e_0^1(t), \cdots, e_0^l(t)$ for $\mathbb{R}^l$ such that

$$M_0(t)e_0^i(t)=0$$

for $i=1,\cdots, l_1$. Make the change of coordinates

$$u^{(0)}=E_0(t)u^{(1)},$$

where $u^{(0)}=(u_1^{(0)},\cdots, u_l^{(0)})$ are the original coordinates, $u^{(1)}=(u_1^{(1)},\cdots, u_l^{(1)})$ are the new coordinates and

$$E_0(t)=(e_0^1(t)\ \vdots\ \cdots\ \vdots\ e_0^l(t)).$$

On the subspace $\mathbb{R}^{l_1}\subseteq\mathbb{R}^l$ spanned by $e_0^1(t),\cdots, e_0^{l_1}(t)$, we continue to change coordinates. Consider the $l_1\times l_1$ matrix

$$M_1(t)=\left(\frac{\partial}{\partial u_i}\frac{d^2}{dt^2}\frac{\partial}{\partial u_j}H(\lambda(t), x(t), u^0(t))\right),$$

where $1\leqq i, j\leqq l_1$. This matrix is symmetric, and by passing to a subinterval if necessary, we can assume it is of rank l_1-l_2 for each $t\in(t^1, t^2)$. Choose an orthonormal basis $e_1^1(t),\cdots, e_1^{l_1}(t)$ for $\mathbb{R}^{l_1}$ such that

$$M_1(t)e_1^i(t)=0$$

for $i=1,\cdots,l_2$. Make the change of coordinates

$$u^{(2)}=E_2(t)u^{(1)},$$

where

$$E_2(t)=\left(\begin{array}{ccc|cc} e_1^1(t) & \cdots & e_1^{l_1}(t) & \multicolumn{2}{c}{0} \\ \hline & & & 1 & 0 \\ & 0 & & & \\ & & & 0 & 1 \end{array}\right)$$

We continue on in this fashion until some $M_k(t)$ is of full rank or it becomes clear that some controls are singular of infinite degree. Then apply the corollary.

Example 6.2. Suppose $u^0(t)$ generates a singular extremal for (2.1) on $[t^1, t^2]$ which is not normal and we wish to apply Theorem 6.2. Once again, to obtain as many necessary conditions as possible, we must make a change of coordinates in the control space.

Since the trajectory is not normal at each $t\in(t^1,t^2)$, there exists more than one linearly independent λ satisfying the constant and linear necessary conditions (6.4) and (6.5). By passing to a subinterval if necessary, we can assume that the dimension of the space of such λ is constant, say ρ, at each $t\in(t^1,t^2)$, and therefore there exist ρ linearly independent solutions $\lambda^1(t),\cdots,\lambda^\rho(t)$ of the adjoint differential equation (2.6) satisfying (6.4) and (6.5).

Define ρ symmetric $l\times l$ matrices:

$$M_0^\sigma(t)=\left(\frac{\partial^2}{\partial u_i\partial u_j}H(\lambda^\sigma(t),x(t),u^0(t))\right)$$

where $1\leqq\sigma\leqq\rho$ and $1\leqq i,j\leqq l$. By passing to a subinterval if necessary we can assume that the rank of the $\rho\cdot l\times l$ matrix,

$$M_0(t)=\begin{pmatrix} M_0^1(t)\\ \vdots \\ M_0^\rho(t)\end{pmatrix},$$

is constant, say $l-l_1$. Choose an orthonormal basis $e_0^1(t),\cdots,e_0^l(t)$ for $\mathbb{R}^l$ such that

$$M_0(t)\,e_0^i(t)=0$$

for $i=1,\cdots,l_1$. Make the change of coordinates

$$u^{(0)}=E_0(t)u^{(1)}$$

where $u^{(0)}$, $u^{(1)}$ and $E_0(t)$ are as in Example 6.1.

On the subspace $\mathbb{R}^{l_1}$ spanned by $e_0^1(t),\cdots,e_0^{l_1}(t)$, we continue to change coordinates. Consider the ρ symmetric $l_1\times l_1$ matrices

$$M_1^\sigma(t)=\left(\frac{\partial}{\partial u_i}\frac{d^2}{dt^2}\frac{\partial}{\partial u_j}H(\lambda^\sigma(t),x(t),u^0(t))\right),$$

where $1 \leqq \sigma \leqq \rho$ and $1 \leqq i, j \leqq l_1$. Define

$$M_1(t) = \begin{pmatrix} M_1^1(t) \\ \vdots \\ M_1^\rho(t) \end{pmatrix}$$

and so on until some $M_k(t)$ is of full rank or it becomes clear that some controls are singular of infinite degree. Then apply Theorem 6.2.

7. Conclusion. The purpose of this paper was to introduce the HMP as a useful tool for constructing high order necessary conditions for optimal control problems with terminal constraints. The HMP is a natural extension of the PMP based on a generalized form of the Pontryagin–Weierstrass condition.

We used the HMP to rigorously demonstrate the GLC for problems with terminal constraints with or without normality. Heretofore the proofs of the GLC relied on a blanket assumption of normality to guarantee their validity.

The HMP can also be used to develop necessary conditions specifically tailored for the problem of interest as in Example 5.3. These special conditions might involve cubic or higher effects of control variations. Further research is needed to discover whether they can be put in a systematic form.

REFERENCES

[1] D. J. BELL, *Singular problem in optimal control—A survey*, Control Systems Centre Report 249, University of Manchester Institute of Science and Technology, Manchester, England, 1974.

[2] R. GABASOV AND F. M. KIRILLOVA, *High order necessary conditions for optimality*, this Journal, 10 (1972), pp. 127–168.

[3] B. S. GOH, *The second variation for the singular Bolza problem*, this Journal, 4 (1966), pp. 309–325.

[4] ———, *Necessary conditions for singular extremals involving multiple control variables*, this Journal, 4 (1966), pp. 716–731.

[5] H. HALKIN, *A maximum principle of the Pontryagin type for systems described by nonlinear difference equations*, this Journal, 4 (1966), pp. 90–111.

[6] N. JACOBSON, *Lie Algebras*, Wiley-Interscience, New York, 1962.

[7] H. J. KELLEY, *A second variation test for singular extremals*, AIAA J., 2 (1964), pp. 1380–1382.

[8] H. J. KELLEY, R. E. KOPP AND H. G. MOYER, *Singular extremals*, Topics in Optimization, G. Leitmann, ed., Academic Press, New York, 1967, pp. 63–101.

[9] R. E. KOPP AND H. G. MOYER, *Necessary conditions for singular extremals*, AIAA J., 3 (1965), pp. 1439–1444.

[10] A. J. KRENER, *A generalization of Chow's theorem and the bang-bang theorem to nonlinear control problems*, this Journal, 12 (1974), pp. 43–52.

[11] ———, *The high order maximal principle*, Geometric Methods in System Theory, D. Q. Mayne and R. W. Brockett, eds., D. Reidel, Dordrecht, Holland, 1973, pp. 174–184.

[12] J. P. MCDANELL AND W. F. POWERS, *Necessary conditions for joining optimal singular and nonsingular subarcs*, this Journal, 9 (1971), pp. 161–173.

[13] H. MAURER, *An example of a continuous junction for a singular control problem of even order*, this Journal, 13 (1975), pp. 899–903.

[14] H. M. ROBBINS, *Optimality of intermediate thrust arcs of rocket trajectories*, AIAA J., 3 (1965). pp. 1094–1098.

[15] ———, *A generalized Legendre–Clebsch condition for the singular cases of optimal control*, IBM J. Res. Develop., 11 (1967), pp. 361–372.

[16] H. J. SUSSMANN, *Minimal realizations of nonlinear systems*, Geometric Methods in Systems Theory, D. Q. Mayne and R. W. Brockett, eds., D. Riedel, Dordrecht, Holland, 1973, pp. 243–252.

[17] K. TAIT, *Singular problems in optimal control*, Ph.D. thesis, Harvard University, Cambridge, MA, 1965.

[18] D. H. JACOBSON, *Totally singular quadratic minimization problems*, IEEE Trans. Automatic Control, AC-16 (1971), pp. 651–658.

[19] H. HERMES AND J. P. LA SALLE, *Functional Analysis and Time Optimal Control*, Academic Press, New York, 1969.

[20] H. HERMES, *High order algebraic conditions for controllability*, Proc. Undine Conference on Algebraic System Theory, 1975.