THE PONTRYAGIN MAXIMUM PRINCIPLE

Héctor J. Sussmann — November 29, 2000

CONTENTS

The purpose of this note is to present a precise statement of a technically simple version of the Pontryagin maximum principle. But before we get there we will add one more example to our list of "missed opportunities," by showing how Carathéodory failed to discover the Hamiltonian maximization condition because of his insistence on using the "wrong" Hamiltonian.

1. Carathéodory's missed opportunity

Constantin Carathéodory (1873-1950) published his work on the calculus of variations in a 1935 book, cf. [**2**]. His discussion of the excess function, and of how Weierstrass' condition can be expressed in Hamiltonian form, is worth quoting in detail[1], because it shows that the author understood the importance of expressing the Weierstrass condition in Hamiltonian form, but was unable to find its simple Hamiltonian meaning because he was using an inappropriate Hamiltonian formalism.

Carathéodory uses x_i for the position variables (that is, our q^i) and y_i for the momenta (that is, our p_i). He works in the neighborhood of a "regular line element" $e = (t, x_i, \dot{x}_i)$, where the "regularity" assumption is the requirement that the Hessian matrix $\{L_{\dot{x}_i,\dot{x}_j}\}_{1\le i,j\le n}$ be positive definite at e. This condition guarantees that, near e, one can express everything in "canonical coordinates" (t, x_i, y_i), i.e., that the map $(t, x, \dot{x}) \to (t, x, y)$ is invertible, if we let $y_i = \chi_i(t, x, \dot{x})$, where $\chi_i(t, x, \dot{x}) = \frac{\partial L}{\partial \dot{x}_i}(t, x, \dot{x})$.

He then expresses the Weierstrass condition, which involves the "line element $(t, x_i, \dot{x}_i)$" together with a nearby "line element" (t, x_i, x_i'), in terms of the corresponding canonical coordinates (t, x_i, y_i) and (t, x_i, y_i'). He uses H to represent what he calls "the Hamiltonian" (that is, our classical Hamiltonian), given by the formula $H = \sum_{i=1}^{n} y_i\dot{x}_i - L(t, x, \dot{x})$, where $\dot{x}_i = \varphi_i(t, x, y)$, and φ is the inverse function of χ (that is, $\varphi(t, x, \chi(t, x, u)) = u$

[1] We follow [**2**], pages 210-212.

and $\chi(t,x,\varphi(t,x,y)) = y$). In addition, he uses H' to represent the Hamiltonian evaluated at the other point (t,x,y') in phase space, that is, $H' = \sum_{i=1}^n y'_i x'_i - L(t,x,x')$, where $x'_i = \varphi_i(t,x,y')$, $y'_i = \chi_i(t,x,x')$.

With these notations, the Weierstrass excess function becomes

$$\begin{aligned}
\mathcal{E}(t,x,\dot{x},x') &= L(t,x,x') - L(t,x,\dot{x}) - \frac{\partial L}{\partial \dot{x}}(t,x,\dot{x})\cdot(x'-\dot{x}) \\
&= L(t,x,x') - L(t,x,\dot{x}) - y\cdot(x'-\dot{x}) \\
&= y\cdot\dot{x} - L(t,x,\dot{x}) - \Big(y\cdot x' - L(t,x,x')\Big) \\
&= y\cdot\dot{x} - L(t,x,\dot{x}) - \Big(y'\cdot x' - L(t,x,x')\Big) - (y-y')\cdot x' \\
&= H - H' - (y-y')\cdot x'\,.
\end{aligned}$$

Since

$$\begin{aligned}
\frac{\partial H'}{\partial y'_j} &= \frac{\partial}{\partial y'_j}\Big(\sum_{i=1}^n y'_i x'_i - L(t,x,x')\Big) \\
&= x'_j + \sum_{i=1}^n y'_i \frac{\partial \varphi_i}{\partial y'_j}(t,x,y') - \frac{\partial}{\partial y'_j}\ Big(L(t,x,\varphi(t,x,y'))\Big) \\
&= x'_j + \sum_{i=1}^n y'_i \frac{\partial \varphi_i}{\partial y'_j}(t,x,y') - \sum_{i=1}^n \frac{\partial L}{\partial x'_i}(t,x,x')\frac{\partial \varphi_i}{\partial y'_j}(t,x,x') \\
&= x'_j + \sum_{i=1}^n y'_i \frac{\partial \varphi_i}{\partial y'_j}(t,x,y') - \sum_{i=1}^n y'_i \frac{\varphi_i}{\partial y'_j}(t,x,x') \\
&= x'_j\,,
\end{aligned}$$

Carathéodory concluded that "the equation

$$\begin{aligned}
\mathcal{E}(t,x,\dot{x},x') &= L' - L - L_{\dot{x}_j}(x'_j - \dot{x}_j) \\
&= H - H' - H'_{y'_j}(y_j - y'_j)
\end{aligned}$$

therefore holds, by which the $\mathcal{E}$ function is represented in the desired form."

2. The maximum principle

The first rigorous statement and proof of the maximum principle appears in the book [**6**]. This "classical" version was then improved by other authors, e.g. L. Cesari [**3**], M. Hestenes [**4**], E. B. Lee and L. Markus [**5**], and L. D. Berkovitz [**1**].

Here we shall present a couple of simple versions of the result, under technical assumptions that will be considerably weakened later.

2.1. A classical version for fixed time interval problems. We begin by quoting a relatively simple version.

We assume that the following conditions are satisfied:

C1. n, m are nonnegative integers;
C2. Q is an open subset of $\mathbb{R}^n$;
C3. U is a closed subset of $\mathbb{R}^m$;
C4. a, b are real numbers such that $a \leq b$;
C5. $Q \times U \times [a,b] \ni (x,u,t) \mapsto f(x,u,t) = (f^1(x,u,t), \ldots, f^n(x,u,t) \in \mathbb{R}^n$ and $Q \times U \times [a,b] \ni (x,u,t) \mapsto L(x,u,t) \in \mathbb{R}$ are continuous maps;
C6. for each $(u,t) \in U \times [a,b]$ the maps

$$Q \ni x \mapsto f(x,u,t) = \Big(f^1(x,u,t), \ldots, f^n(x,u,t)\Big) \in \mathbb{R}^n$$

and

$$Q \ni x \mapsto L(x,u,t) \in \mathbb{R}$$

are continuously differentiable, and their partial derivatives with respect to the x coordinates are continuous functions of (x,u,t);
C7. $\bar{x}$, $\hat{x}$ are given points of Q;
C8. $TCP_{[a,b]}(Q,U,f)$ (the set of all "trajectory-control pairs defined on $[a,b]$ for the data Q,U,f") is the set of all pairs (ξ,η) such that:
   a. $[a,b] \ni t \mapsto \eta(t) \in U$ is a measurable bounded map,
   b. $[a,b] \ni t \mapsto \xi(t) \in Q$ is an absolutely continuous map,
   c. $\dot{\xi}(t) = f(\xi(t),\eta(t),t)$ for almost every $t \in [a,b]$;.
C9. $TCP_{[a,b]}(Q,U,f) \ni (\xi,\eta) \mapsto J(\xi,\eta) \in \mathbb{R}$ is the functional given by

$$J(\xi,\eta) \stackrel{\text{def}}{=} \int_a^b L(\xi(t),\eta(t),t)\,dt\,;$$

C10. $\gamma_* = (\xi_*,\eta_*)$ (the "reference TCP") is such that
   a. $\gamma_* \in TCP_{[a,b]}(Q,U,f)$,
   b. $\xi_*(a) = \bar{x}$ and $\xi_*(b) = \hat{x}$,
   c. $J(\xi_*,\eta_*) \leq J(\xi,\eta)$ for all $(\xi,\eta) \in TCP_{[a,b]}(Q,U,f)$ such that $\xi(a) = \bar{x}$ and $\xi(b) = \hat{x}$.

THEOREM 2.1.1. *Assume that the data n, m, Q, U, a, b, f, L, $\bar{x}$, $\hat{x}$ satisfy conditions C1-C7, $TCP_{[a,b]}(Q,U,f)$ and J are defined by C8-C9, and $\gamma_* = (\xi_*,\eta_*)$ satisfies C10. Define the Hamiltonian H to be the function*

$$Q \times U \times \mathbb{R}^n \times \mathbb{R} \times [a,b] \ni (x,u,p,p_0,t) \mapsto H(x,u,p,p_0,t) \in \mathbb{R}$$

given by

$$H(x,u,p,p_0,t) \stackrel{\text{def}}{=} \langle p, f(x,u,t)\rangle - p_0 L(x,u,t)\,.$$

Then there exists a pair (π,π_0) such that

E1. $[a,b] \ni t \mapsto \pi(t) \in \mathbb{R}^n$ *is an absolutely continuous map;*
E2. $\pi_0 \in \mathbb{R}$ *and* $\pi_0 \geq 0$*;*
E3. $(\pi(t),\pi_0) \neq (0,0)$ *for every* $t \in [a,b]$*;*

E4. *the "adjoint equation" holds, that is,*

$$\dot{\pi}(t) = -\frac{\partial H}{\partial x}(\xi_*(t), \eta_*(t), \pi(t), \pi_0, t)$$

for almost every $t \in [a, b]$;

E5. *the "Hamiltonian maximization condition" holds, that is,*

$$H(\xi_*(t), \eta_*(t), \pi(t), \pi_0, t) = \max\Big\{H(\xi_*(t), u, \pi(t), \pi_0, t) : u \in U\Big\}$$

for almost every $t \in [a, b]$.

Remark 2.1.2. It is clear that

$$\frac{\partial H}{\partial p}(x, u, p, p_0, t) = f(x, u, t)\,.$$

Therefore the adjoint equation, together with the equation of C8-c, say that the pair (ξ_*, π) is a solution of *Hamilton's equations*

$$\text{(2.1)} \qquad \dot{\xi}_*(t) = \frac{\partial H}{\partial p}(\xi_*(t), \eta_*(t), \pi(t), \pi_0, t)\,,$$

$$\text{(2.2)} \qquad \dot{\pi}(t) = -\frac{\partial H}{\partial x}(\xi_*(t), \eta_*(t), \pi(t), \pi_0, t)\,.$$

♢

Remark 2.1.3. The map π is often referred to in the control theory literature as the "adjoint vector." Its proper status from the differential-geometric point of view is that of a *field of covectors* along ξ_*. ♢

Remark 2.1.4. The number π_0 is the "abnormal multiplier." ♢

Remark 2.1.5. A trajectory-control pair for which there exists a pair (π, π_0) that satisfies E1-E5 is called an *extremal.* Therefore the Maximum Principle says that a necessary condition for a TCP to be a solution of the problem of minimizing the functional $J(\xi, \eta)$ subject to the conditions $(\xi, \eta) \in TCP_{[a,b]}(Q, U, f)$ and the endpoint conditions C10-b is that (ξ, η) be an extremal. ♢

Remark 2.1.6. An extremal is *normal* if the pair (π, π_0) can be chosen so that $\pi_0 \neq 0$, and *abnormal* if the pair (π, π_0) can be chosen so that $\pi_0 = 0$, The pair (π, π_0) need not be unique, so in particular it can happen that an extremal is both normal and abnormal. ♢

Remark 2.1.7. Property E3 is referred to as the "nontriviality condition." ♢

Remark 2.1.8. The adjoint equation says that

$$\dot{\pi}(t) = -\pi(t) \cdot \frac{\partial f}{\partial x}(\xi_*(t), \eta_*(t), t) + \pi_0 \frac{\partial L}{\partial x}(\xi_*(t), \eta_*(t), t)\,.$$

In particular, if $(\pi(t), \pi_0) = (0, 0)$ for one value τ of t, then $\pi_0 = 0$, so the equation reduces to

$$\dot{\pi}(t) = -\pi(t) \cdot \frac{\partial f}{\partial x}(\xi_*(t), \eta_*(t), t)\,,$$

which is a linear homogeneous time-varying ordinary differential equation for π. Therefore the fact that $\pi(\tau)=0$ implies that $\pi(t)=0$ for all t, so $(\pi(t),\pi_0)=(0,0)$ for all t. Hence the nontriviality condition could equally well have been stated by saying "for some t" rather than "for all t." $\diamondsuit$

Remark 2.1.9. If we let

$$h(t)=H(\xi_*(t),\eta_*(t),\pi(t),\pi_0,t)\,,$$

then, formally, the derivative of h is given by

$$\begin{aligned}\dot h(t) &= \frac{\partial H}{\partial x}(\Gamma(t))\cdot\dot\xi_*(t)+\frac{\partial H}{\partial u}(\Gamma(t))\cdot\dot\eta_*(t)+\frac{\partial H}{\partial p}(\Gamma(t))\cdot\dot\pi(t)+\frac{\partial H}{\partial t}(\Gamma(t))\\ &= \frac{\partial H}{\partial u}(\Gamma(t))\cdot\dot\eta_*(t)+\frac{\partial H}{\partial t}(\Gamma(t))\,,\end{aligned}$$

where we have written $\Gamma(t)=(\xi_*(t),\eta_*(t),\pi(t),\pi_0,t)$ and used the fact that $\dot\xi_*=\frac{\partial H}{\partial p}$ and $\dot\pi=-\frac{\partial H}{\partial x}$ by Hamilton's equations. If we disregard the fact that $\dot\eta_*(t)$ need not exist (because η_* is not assumed to be diferentiable or even continuous), and in addition we pretend that "$\frac{\partial H}{\partial u}(\xi_*(t),\eta_*(t),\pi(t),\pi_0,t)$ must vanish because the function $u\mapsto H(\xi_*(t),u,\pi(t),\pi_0,t)$ has a maximum at $u=\eta_*(t)$," then we end up with

$$\dot h(t)=\frac{\partial H}{\partial t}(\xi_*(t),\eta_*(t),\pi(t),\pi_0,t)\,.$$

This leads to the conjecture—that we are very far from having proved—that *if the dynamical law f and the Lagrangian L do not depend on t then the function $[a,b]\ni t\mapsto H(\xi_*(t),\eta_*(t),\pi(t),\pi_0,t)$ is constant.* We will see later that this can be made precise and proved rigorously.

2.2. A classical version for variable time interval problems. We assume that the following conditions are satisfied:

- C1. n, m are nonnegative integers;
- C2. Q is an open subset of $\mathbb{R}^n$;
- C3. U is a closed subset of $\mathbb{R}^m$;
- C4. the maps $Q\times U\ni(x,u)\mapsto f(x,u)=(f^1(x,u),\dots,f^n(x,u))\in\mathbb{R}^n$ and $Q\times U\ni(x,u)\mapsto L(x,u)\in\mathbb{R}$ are continuous;
- C5. for each $u\in U$ the maps

  $$Q\ni x\mapsto f(x,u)=\Big(f^1(x,u),\dots,f^n(x,u)\Big)\in\mathbb{R}^n$$

  and $$Q\ni x\mapsto L(x,u)\in\mathbb{R}$$

  are continuously differentiable, and their partial derivatives with respect to the x coordinates are continuous functions of (x,u);
- C6. $\bar x$, $\hat x$ are given points of Q;
- C7. $TCP(Q,U,f)$ (the set of all "trajectory-control pairs for the data Q,U,f") is the set given by

  $$TCP(Q,U,f)\stackrel{\text{def}}{=}\bigcup\Big\{TCP_{[a,b]}(Q,U,f):a,b\in\mathbb{R},a\le b\Big\}\,,$$

where, for $a, b \in \mathbb{R}$ such that $a \leq b$, $TCP_{[a,b]}(Q, U, f)$ is the set of all pairs (ξ, η) such that:

a. $[a, b] \ni t \mapsto \eta(t) \in U$ is a measurable bounded map;
b. $[a, b] \ni t \mapsto \xi(t) \in Q$ is an absolutely continuous map;
c. $\dot{\xi}(t) = f(\xi(t), \eta(t))$ for almost every $t \in [a, b]$.

C8. $TCP(Q, U, f) \ni (\xi, \eta) \mapsto J(\xi, \eta) \in \mathbb{R}$ is the functional given by

$$J(\xi, \eta) \stackrel{\text{def}}{=} \int_a^b L(\xi(t), \eta(t))\, dt \ \text{ for } (\xi, \eta) \in TCP_{[a,b]}(Q, U, f),\ a, b \in \mathbb{R},\ a \leq b.$$

C9. $\gamma_* = (\xi_*, \eta_*)$ (the "reference TCP") are a_*, b_* are such that

a. $a_* \in \mathbb{R}$, $b_* \in \mathbb{R}$, and $\gamma_* \in TCP_{[a_*,b_*]}(Q, U, f)$;
b. $\xi_*(a_*) = \bar{x}$ and $\xi_*(b_*) = \hat{x}$;
c. $J(\xi_*, \eta_*) \leq J(\xi, \eta)$ for all $a, b \in \mathbb{R}$ such that $a \leq b$ and all $(\xi, \eta) \in TCP_{[a,b]}(Q, U, f)$ such that $\xi(a) = \bar{x}$ and $\xi(b) = \hat{x}$.

THEOREM 2.2.1. *Assume that the data n, m, Q, U, f, L, $\bar{x}$, $\hat{x}$ satisfy conditions C1-C6, $TCP(Q, U, f)$ and J are defined by C7-C8, and a_*, b_*, $\gamma_* = (\xi_*, \eta_*)$ satisfy C9. Define the Hamiltonian H to be the function*

$$Q \times U \times \mathbb{R}^n \times \mathbb{R} \ni (x, u, p, p_0) \mapsto H(x, u, p, p_0) \in \mathbb{R}$$

given by

$$H(x, u, p, p_0) \stackrel{\text{def}}{=} \langle p, f(x, u)\rangle - p_0 L(x, u)\,.$$

Then there exists a pair (π, π_0) such that

E1. $[a_*, b_*] \ni t \mapsto \pi(t) \in \mathbb{R}^n$ *is an absolutely continuous map;*
E2. $\pi_0 \in \mathbb{R}$ *and* $\pi_0 \geq 0$*;*
E3. $(\pi(t), \pi_0) \neq (0, 0)$ *for every* $t \in [a_*, b_*]$*;*
E4. *the "adjoint equation" holds, that is,*

$$\dot{\pi}(t) = -\frac{\partial H}{\partial x}(\xi_*(t), \eta_*(t), \pi(t), \pi_0)$$

for almost every $t \in [a_*, b_*]$*;*

E5. *the "Hamiltonian maximization condition with zero value" holds, that is,*

$$0 = H(\xi_*(t), \eta_*(t), \pi(t), \pi_0) = \max\Big\{H(\xi_*(t), u, \pi(t), \pi_0) : u \in U\Big\}$$

for almost every $t \in [a, b]$.

REMARK 2.2.2. If we fix a_* and b_*, and consider only TCPs (ξ, η) belonging to $TCP_{[a_*,b_*]}(Q, U, f)$, then γ_* is obviously a solution of the minimization problem considered in Theorem 2.1.1. So all the conditions given by that theorem should be satisfied here. And, indeed, this is what happens, because *the only difference between the conclusion of Theorem 2.1.1 and that of Theorem 2.2.1 is that in Theorem 2.2.1 we obtain the additional condition that the value of the Hamiltonian must vanish.* ♢

REMARK 2.2.3. We already pointed out in Remark 2.1.9 that, at least formally, the conclusions of Theorem 2.1.1 already imply that the Hamiltonian function $t \mapsto H(\xi_*(t), \eta_*(t), \pi(t), \pi_0)$ is constant, since in our problem f and L do not depend on t. So the only novelty of Theorem 2.2.1 is that the constant now turns out to be equal to zero. $\diamondsuit$

2.3. An example. We consider the "one-dimensional soft landing" problem in which it is desired to find, for a given initial point $(\alpha, \beta) \in \mathbb{R}^2$, a trajectory-control pair (ξ, η) of the system

$$\dot{x} = y\,, \qquad \dot{y} = u\,, \qquad -1 \le u \le 1\,,$$

such that ξ goes from (α, β) to $(0, 0)$ in minimum time.

This is, of course, a problem of the kind discussed in Theorem 2.2.1, so we shall solve it by applying the theorem.

In this case, the configuration space Q is $\mathbb{R}^2$, the control set U is the compact interval $[-1, 1]$, the dynamical law f is given by

$$f(x, y, u) = (y, u)\,,$$

and the Lagrangian is identically equal to 1.

The Hamiltonian H is given by

$$H(x, y, u, p_x, p_y, p_0) = p_x y + p_y u - p_0\,,$$

where we are using p_x, p_y to denote the two components of the momentum variable p.

Assume that (ξ_*, η_*) is a solution of our minimum time problem, and that $(\xi_*, \eta_*) \in TCP_{[a_*, b_*]}(Q, U, f)$.

Then Theorem 2.2.1 tells us that there exists a pair (π, π_0) satisfying all the conditions of the conclusion. Write $\pi(t) = (\pi_x(t), \pi_y(t))$. The adjoint equation then implies

$$\begin{aligned} \dot{\pi}_x(t) &= 0\,, \\ \dot{\pi}_y(t) &= -\pi_x(t)\,. \end{aligned}$$

Therefore the function π_x is constant. (Notice that we are using the fact that π_x is absolutely continuous!) Let $A \in \mathbb{R}$ be such that $\pi_x(t) = A$ for all $t \in [a_*, b_*]$. Then there must exist a constant B such that

$$\pi_y(t) = B - At \quad \text{for} \quad t \in [a_*, b_*]\,.$$

Now, if A and B were both equal to zero, the function π_y would vanish identically and then the Hamiltonian maximization condition would say that the function

$$[-1, 1] \ni u \mapsto 0$$

is maximized by taking $u = \eta_*(t)$, a fact that would give us no information whatsoever about η_*.

Fortunately, the conditions of Theorem 2.2.1 imply that *A and B cannot both vanish.* To see this, observe that if $A = B = 0$ then it follows that

$\pi_x(t) \equiv \pi_y(t) \equiv 0$. But then the nontriviality condition tells us that $\pi_0 \neq 0$. So the value $H(\xi_*(t), \eta_*(t), \pi(t), \pi_0)$ would be equal to $-\pi_0$, which is not equal to zero. This contradicts the fact that, for our time-varying problem, the Hamiltonian is supposed to vanish.

Now that we know that A and B cannot both vanish, there are two possibilities. Either $A = 0$ or $A \neq 0$.

Suppose first that $A = 0$. Then $B \neq 0$, and Hamiltonian maximization tells us that $[-1, 1] \ni u \mapsto Bu$ is maximized by $u = \eta_*(t)$. If $B > 0$, this implies that $\eta_*(t) = 1$ for all t. If $B < 0$ then it follows that $\eta_*(t) = -1$ for all t.

Now assume that $A \neq 0$. Then $t \mapsto B - At \stackrel{\text{def}}{=} \varphi(t)$ is a nonconstant linear function of t. Therefore φ vanishes at most once on the interval $[a_*, b_*]$. If φ never vanishes on $[a_*, b_*]$, or vanishes at one of the endpoints, then $\eta_*(t)$ is either always equal to 1 (if $\varphi(t) > 0$) or always equal to -1 (if $\varphi(t) < 0$). If $\varphi(\tau) = 0$ for some $\tau \in]a_*, b_*[$, then $\eta_*(t)$ will be equal to 1 for $t < \tau$ and to -1 for $t > \tau$ (if φ changes sign at τ from positive to negative) or $\eta_*(t)$ will be equal to -1 for $t < \tau$ and to 1 for $t > \tau$ (if φ changes sign at τ from negative to positive).

So we have proved that η_* is of one of the following four types:

a. constantly equal to 1,
b. constantly equal to -1,
c. constantly equal to -1 for $t < \tau$ and to 1 for $t > \tau$ for some $\tau \in]a_*, b_*[$,
d. constantly equal to 1 for $t < \tau$ and to -1 for $t > \tau$ for some $\tau \in]a_*, b_*[$.

For a problem where the control set U is a compact convex subset of $\mathbb{R}^m$, a control η_* that takes values in the set of extreme points of U is said to be a *bang-bang control*. In our case, we have proved that *all optimal controls are bang-bang and either constant of piecewise constant with one switching.*

Notice that the switchings of the optimal control η_* are determined by the function φ, which in this example happens to be π_y. That is why this function is called the *switching function* for this problem.

Notice also that *we have made essential use of all the conditions given by Theorem 2.2.1.* In particular, *the nontriviality condition was crucial, and the fact that the value of the Hamiltonian is equal to zero was also decisive.*

References

[1] Berkovitz, L. D., *Optimal Control Theory.* Springer-Verlag, New York, 1974.
[2] Carathéodory, C., *Calculus of Variations and Partial Differential Equations of the First Order. Part II: Calculus of Variations.* Eng. transl. by R. B. Dean, 2nd edition, Holden Day, Inc., San Francisco, 1967. (Originally published as *Variationsrechnung und Partielle Differentialgleichungen erster Ordnung.* B. G. Teubner, Berlin, 1935.)
[3] Cesari, L., *Optimization—Theory and Applications.* Springer-Verlag, New York, 1983.
[4] Hestenes, M. R., *Calculus of Variations and Optimal Control Theory.* Wiley, 1966.
[5] Lee, E. B., and L. Markus, *Foundations of Optimal Control Theory.* Wiley, New York, 1967.
[6] Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischenko, *The Mathematical Theory of Optimal Processes.* Wiley, New York, 1962.