2005 WSEAS Int. Conf. on DYNAMICAL SYSTEMS and CONTROL, Venice, Italy, November 2-4, 2005 (pp477-482)

Mixed Equality Constraints in Optimal Control

JAVIER F ROSENBLUETH
Department of Mathematical and Numerical Methods
National Autonomous University of Mexico
IIMAS–UNAM, Apartado Postal 20-726, México DF 01000
MEXICO
jfrl@servidor.unam.mx

Abstract: This paper concerns a derivation of second order necessary conditions for a fixed-endpoint control problem of Lagrange involving mixed equality constraints. These conditions are obtained by means of an implicit function theorem approach applied to the corresponding conditions for a control problem without constraints.

Key-Words: Optimal control, mixed equality constraints, second order necessary conditions, normality.

1 Introduction

According to Gilbert and Bernstein [1], mathematical rigorous treatments of second order necessary conditions for problems in optimal control seem to be limited. In their paper, two widely quoted works due to Hestenes [2] and Warga [3] are mentioned and the main results obtained in [1] are compared with those references.

Hestenes, whose work is the earliest, considered a fairly general optimal control problem but, quoting [1], made the standard assumption that the control set is open. The aim of this paper is to show how the necessary conditions obtained in [2] for such an unconstrained problem can be used to derive such conditions for problems with mixed equality constraints. The approach is based on a uniform implicit function theorem established precisely in [2].

Let us begin by stating the problem we shall be concerned with together with well-known first order necessary conditions. The results of this section are taken up from [2].

Suppose we are given an interval $T := [t_0, t_1]$ in $\mathbf{R}$, two points ξ_0, ξ_1 in $\mathbf{R}^n$, a set $\mathcal{A}$ in $T \times \mathbf{R}^n \times \mathbf{R}^m$, and functions L and f mapping $T \times \mathbf{R}^n \times \mathbf{R}^m$ to $\mathbf{R}$ and $\mathbf{R}^n$ respectively.

Denote by X the space of piecewise C^1 functions mapping T to $\mathbf{R}^n$, by $\mathcal{U}$ the space of piecewise continuous functions mapping T to $\mathbf{R}^m$, set $Z := X \times \mathcal{U}$,

$$Z(\mathcal{A}) := \{(x,u) \in Z \mid (t, x(t), u(t)) \in \mathcal{A} \ (t \in T)\},$$

$$D(f) := \{(x,u) \in Z \mid \dot{x}(t) = f(t, x(t), u(t)) \quad (t \in T)\},$$

$$Z_e(\mathcal{A}, f) := \{(x,u) \in Z(\mathcal{A}) \cap D(f) \mid x(t_0) = \xi_0, \ x(t_1) = \xi_1\},$$

and consider the functional $I: Z \to \mathbf{R}$ given by $I(x,u) := \int_{t_0}^{t_1} L(t, x(t), u(t))dt \ ((x,u) \in Z)$. The problem we shall deal with, which we label $\mathrm{P}(\mathcal{A}, f, I)$, is that of minimizing I over $Z_e(\mathcal{A}, f)$.

Elements of Z will be called *processes*, and a process (x,u) *solves* $\mathrm{P}(\mathcal{A}, f, I)$ if $(x,u) \in Z_e(\mathcal{A}, f)$ and $I(x,u) \leq I(y,v)$ for all $(y,v) \in Z_e(\mathcal{A}, f)$.

For any $(x,u) \in Z$ we shall use the notation $(\tilde{x}(t))$ to represent $(t, x(t), u(t))$, and '$*$' denotes transpose.

Assumptions

A. The functions L, f and their partial derivatives with respect to x are continuous, and $\mathcal{A}$ is *admissible* in the sense that, for each $(s,y,v) \in \mathcal{A}$, there exist and $u: [s-\delta, s+\delta] \to \mathbf{R}^m$ continuous such that $u(s) = v$ and $(t, x, u(t)) \in \mathcal{A}$ for all $(t,x) \in T \times \mathbf{R}^n$ such that $|(t,x) - (s,y)| < \delta$.

B. The functions L, f are C^2 and $\mathcal{A}$ is (relatively) open.

For all (t,x,u,p,λ) in $T \times \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n \times \mathbf{R}$ let $H(t,x,u,p,\lambda) := \langle p, f(t,x,u)\rangle - \lambda L(t,x,u)$.

1.1 Theorem: *Assume* (A) *holds and* (x_0, u_0) *solves* $\mathrm{P}(\mathcal{A}, f, I)$. *Then there exist* $\lambda_0 \geq 0$ *and* $p \in X$, *not both zero, such that*

a. $\dot{p}(t) = -H_x^*(\tilde{x}_0(t), p(t), \lambda_0)$ *on every interval of continuity of* u_0.

b. $H(t, x_0(t), u, p(t), \lambda_0) \leq H(\tilde{x}_0(t), p(t), \lambda_0)$ *for all* $(t, u) \in T \times \mathbf{R}^m$ *with* $(t, x_0(t), u) \in \mathcal{A}$.

1.2 Note: In Theorem 1.1, the function $t \mapsto H(\tilde{x}_0(t), p(t), \lambda_0)$ is continuous on T. Also, if (B) holds, then $H_u(\tilde{x}_0(t), p(t), \lambda_0) = 0$ and $H_{uu}(\tilde{x}_0(t), p(t), \lambda_0) \leq 0$ $(t \in T)$.

In the remaining of this section we assume that (B) holds.

1.3 Definitions:

• For all $(x, u) \in Z$ the set of *admissible variations along* (x, u) will be given by

$$Y_0(x, u) := \{(y, v) \in Z \mid y(t_0) = y(t_1) = 0,$$

$$\dot{y}(t) = A(t)y(t) + B(t)v(t) \ (t \in T)\},$$

where $A(t) := f_x(\tilde{x}(t))$, $B(t) := f_u(\tilde{x}(t))$ $(t \in T)$.

• A process (x, u) will be said to be *normal* to $\mathrm{P}(\mathcal{A}, f, I)$ if the equations

$$\begin{aligned} \dot{p}(t) &= -A^*(t)p(t) \quad [\, = -H_x^*(\tilde{x}(t), p(t), 0) \,] \\ 0 &= B^*(t)p(t) \quad [\, = H_u^*(\tilde{x}(t), p(t), 0) \,] \end{aligned}$$

have no nonnull solution p on T.

• For all $(x, u) \in Z$ let $M_0(x, u)$ denote the set of all $(\lambda_0, p) \in \mathbf{R} \times X$ such that $\lambda_0 \geq 0$, $\lambda_0 + |p| \neq 0$, $\dot{p}(t) = -H_x^*(t, \lambda_0)$, and $H_u(t, \lambda_0) = 0$ $(t \in T)$, where $H(t, \lambda_0)$ denotes $H(\tilde{x}(t), p(t), \lambda_0)$, and consider the following sets:

$$\mathcal{E}_0 := \{(x, u, p) \in Z \times X \mid (x, u) \in D(f)$$

$$\text{and } (1, p) \in M_0(x, u)\}$$

$$\mathcal{H}_0 := \{(x, u, p) \in Z \times X \mid J_0((x, u); (y, v)) \geq 0$$

$$\text{for all } (y, v) \in Y_0(x, u)\}$$

where, for all $(y, v) \in Z$,

$$J_0((x, u); (y, v)) = \int_{t_0}^{t_1} 2\Omega_0(t, y(t), v(t))dt$$

and, for all $(t, y, v) \in T \times \mathbf{R}^n \times \mathbf{R}^m$,

$$2\Omega_0(t, y, v) := -\big[\langle y, H_{xx}(t, 1)y\rangle +$$

$$2\langle y, H_{xu}(t, 1)v\rangle + \langle v, H_{uu}(t, 1)v\rangle\big]$$

where $H(t, 1)$ denotes $H(\tilde{x}(t), p(t), 1)$.

1.4 Lemma: *Suppose* $(x_0, u_0) \in Z_e(\mathcal{A}, f)$ *is normal to* $\mathrm{P}(\mathcal{A}, f, I)$ *and* $(y, v) \in Y_0(x_0, u_0)$. *Then there exist* $\delta > 0$ *and a one-parameter family* $(x(\cdot, \epsilon), u(\cdot, \epsilon)) \in Z_e(\mathcal{A}, f)$ $(|\epsilon| < \delta)$ *such that*

i. $x(t, 0) = x_0(t)$, $u(t, 0) = u_0(t)$ $(t \in T)$.

ii. $x_\epsilon(t, 0) = y(t)$, $u_\epsilon(t, 0) = v(t)$ $(t \in T)$.

The following result summarizes in a succinct way first and second order necessary conditions for problem $\mathrm{P}(\mathcal{A}, f, I)$ under assumption (B). We provide a simple proof, similar to the one given in [2], based on the results quoted above.

1.5 Theorem: *Suppose* (x, u) *solves* $\mathrm{P}(\mathcal{A}, f, I)$. *Then* $M_0(x, u) \neq \emptyset$. *If* (x, u) *is normal to* $\mathrm{P}(\mathcal{A}, f, I)$ *then there exists a unique* $p \in X$ *such that* $(x, u, p) \in \mathcal{E}_0$. *Moreover,* $(x, u, p) \in \mathcal{H}_0$.

Proof: Suppose (x_0, u_0) solves $\mathrm{P}(\mathcal{A}, f, I)$. By Theorem 1.1 and Note 1.2, $M_0(x_0, u_0) \neq \emptyset$. Let $(\lambda_0, p) \in M_0(x_0, u_0)$ and suppose (x_0, u_0) is normal to $\mathrm{P}(\mathcal{A}, f, I)$. This implies that $\lambda_0 \neq 0$ and, if $(\lambda_0, q) \in M_0(x_0, u_0)$, then $r := p - q$ satisfies $\dot{r}(t) = -A^*(t)r(t)$, $0 = B^*(t)r(t)$ $(t \in T)$, implying that $p \equiv q$. Clearly we can choose $\lambda_0 = 1$ since $(1, p/\lambda_0) \in M_0(x_0, u_0)$. Suppose therefore that $(x_0, u_0, p) \in \mathcal{E}_0$. To show that $(x_0, u_0, p) \in \mathcal{H}_0$, define

$$K(x, u) := \langle p(t_1), \xi_1\rangle - \langle p(t_0), \xi_0\rangle +$$

$$\int_{t_0}^{t_1} F(t, x(t), u(t))dt \qquad ((x, u) \in Z)$$

where, for all $(t, x, u) \in T \times \mathbf{R}^n \times \mathbf{R}^m$,

$$F(t, x, u) = L(t, x, u) - \langle p(t), f(t, x, u)\rangle - \langle \dot{p}(t), x\rangle.$$

Observe that

$$F(t, x, u) = -H(t, x, u, p(t), 1) - \langle \dot{p}(t), x\rangle$$

and, if $(x, u) \in Z_e(\mathcal{A}, f)$, then $K(x, u) = I(x, u)$. Let $(y, v) \in Y_0(x_0, u_0)$ and let $(x(\cdot, \epsilon), u(\cdot, \epsilon)) \in Z_e(\mathcal{A}, f)$ $(|\epsilon| < \delta)$ be a one-parameter family satisfying Lemma 1.4. Hence

$$g(\epsilon) := K(x(\cdot, \epsilon), u(\cdot, \epsilon)) = I(x(\cdot, \epsilon), u(\cdot, \epsilon))$$

satisfies $g(\epsilon) \geq g(0) = K(x_0, u_0) = I(x_0, u_0)$ for all $|\epsilon| < \delta$. Note that $F_x(\tilde{x}_0(t)) = 0$ and $F_u(\tilde{x}_0(t)) = 0$ and, therefore,

$$\begin{aligned} 0 &\leq g''(0) = K''((x_0, u_0); (y, v)) \\ &= J_0((x_0, u_0); (y, v)). \ \blacksquare \end{aligned}$$

2 Auxiliary results

In this section we state three results (see [2] for details) which play a fundamental role in the deriviation of second order conditions for problem

$P(\mathcal{A}, f, I)$ when $\mathcal{A}$ is defined in terms of mixed equality constraints.

2.1 Theorem (Implicit function): *Let $f: A \to \mathbf{R}^n$ where $A \subset \mathbf{R}^m \times \mathbf{R}^n$ is open. Suppose f and $f_x(t,x)$ are continuous on A and there exist $T_0 \subset \mathbf{R}^m$ compact and $x_0: T_0 \to \mathbf{R}^n$ continuous such that, for all $t \in T_0$,*

i. $(t, x_0(t)) \in A$.
ii. $f(t, x_0(t)) = 0$ *and* $|f_x(t, x_0(t))| \neq 0$.

Then there exist T neighborhood of T_0, $\epsilon > 0$ and $x: T \to \mathbf{R}^n$ continuous, such that

a. $x(t) = x_0(t)$ *for all* $t \in T_0$.
b. $f(t, x(t)) = 0$ *for all* $t \in T$.
c. $(t \in T,\ f(t,x) = 0$ *and* $|x - x(t)| < \epsilon) \Rightarrow x = x(t)$.
d. $f \in C^m(A) \Rightarrow x \in C^m(T)$.

2.2 Lemma: *Let φ be a C^1 function mapping $\mathbf{R}^n \times \mathbf{R}^m$ to $\mathbf{R}^q$. Let*

$$S := \{(x,u) \in \mathbf{R}^n \times \mathbf{R}^m \mid \varphi(x,u) = 0\}$$

and suppose the matrix $\varphi_u(x,u)$ has rank q on S. Then there exist a neighborhood $\tilde{S}$ of S and a continuous function $U: \tilde{S} \to \mathbf{R}^m$ such that

a. $(x, U(x,u)) \in S$ *for all* $(x,u) \in \tilde{S}$.
b. $U(x,u) = u$ *for all* $(x,u) \in S$.

If φ is of class C^r then U can be chosen to be of class C^r on $\tilde{S}$.

Proof: Let $h: \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^q \to \mathbf{R}^q$ be given by $h(x,u,b) := \varphi(x, u + \varphi_u^*(x,u)b)$. We have h and h_b continuous on $\mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^q$ and, for all $(x,u) \in S$, $h(x,u,0) = 0$ and

$$|h_b(x,u,0)| = |\varphi_u(x,u)\varphi_u^*(x,u)| \neq 0.$$

Suppose first that $S \subset \mathbf{R}^n \times \mathbf{R}^m$ is compact. By Theorem 2.1 there exist $\tilde{S}$ neighborhood of S, $\epsilon > 0$ and $B: \tilde{S} \to \mathbf{R}^q$ continuous, such that

a. $B(x,u) = 0$ for all $(x,u) \in S$.
b. $h(x,u,B(x,u)) = 0$ for all $(x,u) \in \tilde{S}$.
c. $((x,u) \in \tilde{S},\ h(x,u,b) = 0$ and $|b - B(x,u)| < \epsilon) \Rightarrow b = B(x,u)$.
d. $h \in C^r \Rightarrow B \in C^r(\tilde{S})$.

Let $U(x,u) = u + \varphi_u^*(x,u)B(x,u)$ for all $(x,u) \in \tilde{S}$. Then U satisfies the required properties. For the case S not compact and the fact that U can be chosen to be of class C^r on $\tilde{S}$, we refer to Hestenes' book [2]. ∎

2.3 Note: Observe that the function $B: \tilde{S} \to \mathbf{R}^q$ satisfies

$$B'(x,u) = -(h_b(x,u,0))^{-1} h_{x,u}(x,u,0)$$

for all $(x,u) \in S$ and therefore, if $D(x,u) := [\varphi_u(x,u)\varphi_u^*(x,u)]^{-1}$, then

$$B_x(x,u) = -D(x,u)\varphi_x(x,u),$$

$$B_u(x,u) = -D(x,u)\varphi_u(x,u)$$

for all $(x,u) \in S$. Hence $U: \tilde{S} \to \mathbf{R}^m$ defined by $U(x,u) = u + \varphi_u^*(x,u)B(x,u)$ satisfies, for any $(x,u) \in S$,

$$U_x(x,u) = -\varphi_u^*(x,u)D(x,u)\varphi_x(x,u),$$

$$U_u(x,u) = I - \varphi_u^*(x,u)D(x,u)\varphi_u(x,u).$$

To state the last auxiliary result we shall need, suppose we are given an interval $T = [a,b]$ in $\mathbf{R}$ and functions f and g mapping $T \times \mathbf{R}^m$ to $\mathbf{R}$ and $\mathbf{R}^q$ respectively, with f, g continuous and having continuous partial derivatives with respect to u. Let

$$S := \{(t,u) \in T \times \mathbf{R}^m \mid g(t,u) = 0\}$$

and suppose the matrix $g_u(t,u)$ has rank q on S. Denote by $\mathcal{U}$, as in Section 1, the space of piecewise continuous functions mapping T to $\mathbf{R}^m$ and let

$$\mathcal{U}(S) := \{u \in \mathcal{U} \mid (t, u(t)) \in S\ (t \in T)\}.$$

2.4 Lemma: *Suppose $u_0 \in \mathcal{U}(S)$ is such that $f(t, u_0(t)) \leq f(t,u)$ for all $t \in T$ with $(t,u) \in S$. Then there exists a unique $\mu: T \to \mathbf{R}^q$ such that, if*

$$F(t,u) := f(t,u) + \langle \mu(t), g(t,u) \rangle$$

for all $(t,u) \in T \times \mathbf{R}^m$), then $F_u(t, u_0(t)) = 0$ $(t \in T)$. Moreover, $\langle h, F_{uu}(t, u_0(t))h \rangle \geq 0$ for all $h \in \mathbf{R}^m$ such that $g_u(t, u_0(t))h = 0$. The function μ is piecewise continuous on T and continuous at each point of continuity of u_0.

3 Mixed equality constraints

Consider problem $P(\mathcal{A}, f, I)$ of Section 1. Assume L, f are C^2 and

$$\mathcal{A} = \{(t,x,u) \in T \times \mathbf{R}^n \times \mathbf{R}^m \mid \varphi(t,x,u) = 0\}$$

where $\varphi: T \times \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^q$ is of class C^2 and the matrix $\varphi_u(t,x,u)$ has rank q on $\mathcal{A}$. Denote by $\mathcal{U}_q$ the space of piecewise continuous functions mapping T to $\mathbf{R}^q$.

For all (t,x,u,p,μ,λ) in $T \times \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^n \times \mathbf{R}^q \times \mathbf{R}$ let

$$H(t,x,u,p,\mu,\lambda) := \langle p, f(t,x,u) \rangle -$$

$$\lambda L(t,x,u) - \langle \mu, \varphi(t,x,u)\rangle.$$

Let us begin by showing how first order necessary conditions for this constrained problem can be derived from those for the unconstained one stated in Section 1.

3.1 Theorem: *If (x_0, u_0) solves $\mathrm{P}(\mathcal{A}, f, I)$, then there exist $\lambda_0 \geq 0$, $p \in X$, and $\mu \in \mathcal{U}_q$ continuous on each interval of continuity of u_0, not vanishing simultaneously on T, such that*

a. *$\dot{p}(t) = -H_x^*(t,\lambda_0)$ and $H_u(t,\lambda_0) = 0$ on every interval of continuity of u_0.*

b. *$H(t, x_0(t), u, p(t), \mu(t), \lambda_0) \leq H(t, \lambda_0)$ for all $(t,u) \in T \times \mathbf{R}^m$ with $(t, x_0(t), u) \in \mathcal{A}$*

where $H(t,\lambda_0)$ denotes $H(\tilde{x}_0(t), p(t), \mu(t), \lambda_0)$.

Proof: By Lemma 2.2 there exist a neighborhood $\mathcal{B}$ of $\mathcal{A}$ and a C^2 function $U\colon \mathcal{B} \to \mathbf{R}^m$ such that $(t, x, U(t,x,u)) \in \mathcal{A}$ for all $(t,x,u) \in \mathcal{B}$, and $U(t,x,u) = u$ for all $(t,x,u) \in \mathcal{A}$. Set

$$\hat{f}(t,x,u) := f(t,x,U(t,x,u)),$$

$$\hat{L}(t,x,u) := L(t,x,U(t,x,u))$$

for all $(t,x,u) \in \mathcal{B}$, and

$$\hat{I}(x,u) := \int_{t_0}^{t_1} \hat{L}(t, x(t), u(t))dt$$

for all $(x,u) \in Z(\mathcal{B})$. Let us prove that (x_0, u_0) solves $\mathrm{P}(\mathcal{B}, \hat{f}, \hat{I})$, i.e.,

i. $(x_0, u_0) \in Z_e(\mathcal{B}, \hat{f})$.

ii. $\hat{I}(x_0, u_0) \leq \hat{I}(\hat{x}, \hat{u})$ for all $(\hat{x}, \hat{u}) \in Z_e(\mathcal{B}, \hat{f})$.

Since (x_0, u_0) solves $\mathrm{P}(\mathcal{A}, f, I)$ we have $(x_0, u_0) \in Z_e(\mathcal{A}, f)$ and $I(x_0, u_0) \leq I(x,u)$ for all $(x,u) \in Z_e(\mathcal{A}, f)$. Hence $(t, x_0(t), u_0(t)) \in \mathcal{A} \subset \mathcal{B}$ $(t \in T)$, $\dot{x}_0(t) = \hat{f}(t, x_0(t), u_0(t))$, and so (i) holds. To prove (ii) let $(\hat{x}, \hat{u}) \in Z_e(\mathcal{B}, \hat{f})$ and define

$$x(t) := \hat{x}(t),\ \ u(t) := U(t, \hat{x}(t), \hat{u}(t)) \quad (t \in T).$$

Then $(t, x(t), u(t)) \in \mathcal{A}$, $\dot{x}(t) = f(t, x(t), u(t))$ $(t \in T)$, and so $(x,u) \in Z_e(\mathcal{A}, f)$. Also $I(x,u) = \hat{I}(\hat{x}, \hat{u})$ and $\hat{I}(x_0, u_0) = I(x_0, u_0)$ and so (ii) holds.

Now, by Theorem 1.1 applied to (x_0, u_0) with respect to $\mathrm{P}(\mathcal{B}, \hat{f}, \hat{I})$, there exist $\lambda_0 \geq 0$ and $p \in X$, not both zero, such that if

$$\hat{H}(t,x,u) := \langle p(t), \hat{f}(t,x,u)\rangle - \lambda_0 \hat{L}(t,x,u)$$

for all $(t,x,u) \in \mathcal{B}$, then

a. $\dot{p}(t) = -\hat{H}_x^*(\tilde{x}_0(t))$ on every interval of continuity of u_0.

b. $\hat{H}(t, x_0(t), u) \leq \hat{H}(\tilde{x}_0(t))$ for all $(t,u) \in T \times \mathbf{R}^m$ with $(t, x_0(t), u) \in \mathcal{B}$.

Define

$$G(t,x,u) := \langle p(t), f(t,x,u)\rangle - \lambda_0 L(t,x,u)$$

so that $\hat{H}(t,x,u) = G(t,x,U(t,x,u))$ for all $(t,x,u) \in \mathcal{B}$. Since $U(t,x,u) = u$ for all $(t,x,u) \in \mathcal{A}$, it follows from (b) that

$$G(t, x_0(t), u) \leq G(\tilde{x}_0(t))$$

for all $(t,u) \in T \times \mathbf{R}^m$ with $(t, x_0(t), u) \in \mathcal{A}$. Let $g(t,u) := \varphi(t, x_0(t), u)$ and $h(t,u) := -G(t, x_0(t), u)$ for all $(t,u) \in T \times \mathbf{R}^m$ and set

$$S := \{(t,u) \in T \times \mathbf{R}^m \mid g(t,u) = 0\}.$$

Since $h(t, u_0(t)) \leq h(t,u)$ for all $(t,u) \in S$, by Lemma 2.4 there exists a unique $\mu \in \mathcal{U}_q$ such that, if we set

$$F(t,u) := h(t,u) + \langle \mu(t), g(t,u)\rangle$$

then $F_u(t, u_0(t)) = 0$. In other words, if

$$\hat{G}(t,x,u) := G(t,x,u) - \langle \mu(t), \varphi(t,x,u)\rangle,$$

so that $\hat{G}(t, x_0(t), u) = -F(t,u)$, then

$$0 = \hat{G}_u(\tilde{x}_0(t)) = G_u(\tilde{x}_0(t)) - \mu^*(t)\varphi_u(\tilde{x}_0(t)).$$

The function μ is continuous on each interval of continuity of u_0. Since, for all $(t,x,u) \in \mathcal{B}$, we have $(t, x, U(t,x,u)) \in \mathcal{A}$ and $\hat{H}(t,x,u) = G(t,x,U(t,x,u))$, it follows that

$$\hat{H}(t,x,u) = \hat{G}(t,x,U(t,x,u)) \qquad ((t,x,u) \in \mathcal{B}).$$

Consequently,

$$\begin{aligned}\hat{H}_x(\tilde{x}_0(t)) &= \hat{G}_x(\tilde{x}_0(t)) + \hat{G}_u(\tilde{x}_0(t))U_x(\tilde{x}_0(t)) \\ &= \hat{G}_x(\tilde{x}_0(t)).\end{aligned}$$

Since $\hat{G}(t,x,u) = H(t,x,u,p(t),\mu(t),\lambda_0)$, the conditions of the theorem hold. ∎

3.2 Corollary: *If (x_0, u_0) solves $\mathrm{P}(\mathcal{A}, f, I)$ and (p, μ, λ_0) is as in Theorem 3.1, then*

$$\langle h, H_{uu}(\tilde{x}_0(t), p(t), \mu(t), \lambda_0)h\rangle \leq 0$$

for all $h \in \mathbf{R}^m$ such that $\varphi_u(\tilde{x}_0(t))h = 0$.

3.3 Note: Suppose that, for some $(x,u) \in Z$, $p \in X$, $\lambda \in \mathbf{R}$ and a function $\mu\colon T \to \mathbf{R}^q$, we have

$$0 = H_u(\tilde{x}(t), p(t), \mu(t), \lambda)$$

$$[= p^*(t)f_u(\tilde{x}(t)) - \lambda L_u(\tilde{x}(t)) - \mu^*(t)\varphi_u(\tilde{x}(t))]$$

for all $t \in T$. Then $\mu \in \mathcal{U}_q$ and

$$\mu^*(t) = \{p^*(t) f_u(\tilde{x}(t)) - \lambda L_u(\tilde{x}(t))\} \varphi_u^*(\tilde{x}(t)) D(t)$$

for all $t \in T$, where $D(t) = [\varphi_u(\tilde{x}(t)) \varphi_u^*(\tilde{x}(t))]^{-1}$.

3.4 Definitions:

- For all $(x, u) \in Z$ define the set $Y_1(x, u)$ of *admissible variations along* (x, u) as the set of all $(y, v) \in Z$ satisfying

i. $y(t_0) = y(t_1) = 0$;
ii. $\dot{y}(t) = A(t) y(t) + B(t) v(t)$ $(t \in T)$;
iii. $\varphi_x(\tilde{x}(t)) y(t) + \varphi_u(\tilde{x}(t)) v(t) = 0$ $(t \in T)$,

where $A(t) := f_x(\tilde{x}(t))$, $B(t) := f_u(\tilde{x}(t))$ $(t \in T)$.

- For all $(x, u) \in Z$ let $M_1(x, u)$ denote the set of all $(\lambda_0, p, \mu) \in \mathbf{R} \times X \times \mathcal{U}_q$ such that $\lambda_0 \geq 0$, $\lambda_0 + |p| + |\mu| \neq 0$, $\dot{p}(t) = -H_x^*(t, \lambda_0)$, and $H_u(t, \lambda_0) = 0$ $(t \in T)$, where $H(t, \lambda_0)$ denotes $H(\tilde{x}(t), p(t), \mu(t), \lambda_0)$, and consider the following sets:

$$\mathcal{E}_1 := \{(x, u, p, \mu) \in Z \times X \times \mathcal{U}_q \mid (x, u) \in D(f) \text{ and } (1, p, \mu) \in M_1(x, u)\}$$

$$\mathcal{H}_1 := \{(x, u, p, \mu) \in Z \times X \times \mathcal{U}_q \mid J_1((x, u); (y, v)) \geq 0 \text{ for all } (y, v) \in Y_1(x, u)\}$$

where, for all $(y, v) \in Z$,

$$J_1((x, u); (y, v)) = \int_{t_0}^{t_1} 2\Omega_1(t, y(t), v(t)) dt$$

and, for all $(t, y, v) \in T \times \mathbf{R}^n \times \mathbf{R}^m$,

$$2\Omega_1(t, y, v) := -[\langle y, H_{xx}(t, 1) y\rangle + 2\langle y, H_{xu}(t, 1) v\rangle + \langle v, H_{uu}(t, 1) v\rangle]$$

where $H(t, 1)$ denotes $H(\tilde{x}(t), p(t), \mu(t), 1)$.

- A process (x, u) will be said to be *normal* to $\mathrm{P}(\mathcal{A}, f, I)$ if, given $(p, \mu) \in X \times \mathcal{U}_q$ such that, for all $t \in T$,

$$\dot{p}(t) = -A^*(t) p(t) + \varphi_x^*(\tilde{x}(t)) \mu(t)$$
$$[\, = -H_x^*(\tilde{x}(t), p(t), \mu(t), 0)\,]$$
$$0 = B^*(t) p(t) - \varphi_u^*(\tilde{x}(t)) \mu(t)$$
$$[\, = H_u^*(\tilde{x}(t), p(t), \mu(t), 0)\,]$$

then $p \equiv 0$. In this event, by Note 3.3, also $\mu \equiv 0$. Note that, if (x, u) is normal to $\mathrm{P}(\mathcal{A}, f, I)$ and $(\lambda_0, p, \mu) \in M_1(x, u)$, then $\lambda_0 \neq 0$.

3.5 Note: Let $(x, u) \in Z$ and suppose that, given $(p, \mu) \in X \times \mathcal{U}_q$ such that, for all $t \in T$,

i. $\dot{p}(t) = -A^*(t) p(t) + \varphi_x^*(\tilde{x}(t)) \mu(t)$;
ii. $p^*(t) B(t) h = 0$ for all $h \in \mathbf{R}^m$ satisfying $\varphi_u(\tilde{x}(t)) h = 0$,

necessarily $p \equiv 0$. Then (x, u) is normal to $\mathrm{P}(\mathcal{A}, f, I)$.

3.6 Lemma: *Consider problem* $\mathrm{P}(\mathcal{B}, \hat{f}, \hat{I})$ *given in the proof of Theorem 3.1 and suppose* (x_0, u_0) *is normal to* $\mathrm{P}(\mathcal{A}, f, I)$. *Then* (x_0, u_0) *is normal to* $\mathrm{P}(\mathcal{B}, \hat{f}, \hat{I})$.

Proof: We want to prove that $p \equiv 0$ is the only solution of

$$\dot{p}(t) = -\hat{A}^*(t) p(t), \quad 0 = \hat{B}^*(t) p(t) \qquad (t \in T)$$

where

$$\hat{A}(t) = \hat{f}_x(\tilde{x}_0(t)) = A(t) + B(t) U_x(\tilde{x}_0(t)),$$

$$\hat{B}(t) = \hat{f}_u(\tilde{x}_0(t)) = B(t) U_u(\tilde{x}_0(t)).$$

This follows since, by Note 2.3, the above relations correspond to

$$\begin{aligned} \dot{p}(t) &= -A^*(t) p(t) - U_x^*(\tilde{x}_0(t)) B^*(t) p(t) \\ &= -A^*(t) p(t) + \varphi_x^*(\tilde{x}_0(t)) \mu(t) \end{aligned}$$

and

$$\begin{aligned} 0 &= U_u^*(\tilde{x}_0(t)) B^*(t) p(t) \\ &= B^*(t) p(t) - \varphi_u^*(\tilde{x}_0(t)) \mu(t) \end{aligned}$$

where $\mu(t)$ is equal to

$$[\varphi_u(\tilde{x}_0(t)) \varphi_u^*(\tilde{x}_0(t))]^{-1} \varphi_u(\tilde{x}_0(t)) B^*(t) p(t).$$

Thus the normality of (x_0, u_0) to $\mathrm{P}(\mathcal{A}, f, I)$ implies that $p \equiv 0$. ∎

We are now in a position to establish first and second order conditions for the problem with mixed equality constraints.

3.7 Theorem: *Suppose* (x, u) *solves* $\mathrm{P}(\mathcal{A}, f, I)$. *Then* $M_1(x, u) \neq \emptyset$. *If* (x, u) *is normal to* $\mathrm{P}(\mathcal{A}, f, I)$ *then there exists a unique* $(p, \mu) \in X \times \mathcal{U}_q$ *such that* $(x, u, p, \mu) \in \mathcal{E}_1$. *Moreover,* $(x, u, p, \mu) \in \mathcal{H}_1$.

Proof: Suppose (x_0, u_0) solves $\mathrm{P}(\mathcal{A}, f, I)$. By Theorem 3.1, $M_1(x_0, u_0) \neq \emptyset$. Let $(\lambda_0, p, \mu) \in M_1(x_0, u_0)$ and suppose (x_0, u_0) is normal to $\mathrm{P}(\mathcal{A}, f, I)$. This implies that $\lambda_0 \neq 0$ and, if $(\lambda_0, q, \nu) \in M_1(x_0, u_0)$ then $r := p - q$ satisfies

$$\dot{r}(t) = -A^*(t) r(t) + \varphi_x^*(\tilde{x}_0(t))[\mu(t) - \nu(t)]$$

$$0 = B^*(t) r(t) - \varphi_u^*(\tilde{x}_0(t))[\mu(t) - \nu(t)]$$

implying that $p \equiv q$ and $\mu \equiv \nu$.

Let $(p, \mu) \in X \times \mathcal{U}_q$ be the unique pair such that $(x_0, u_0, p, \mu) \in \mathcal{E}_1$ and define

$$K(x,u) := \langle p(t_1), \xi_1 \rangle - \langle p(t_0), \xi_0 \rangle + \int_{t_0}^{t_1} F(t, x(t), u(t)) dt \quad ((x,u) \in Z)$$

where, for all $(t, x, u) \in T \times \mathbf{R}^n \times \mathbf{R}^m$,

$$F(t,x,u) := L(t,x,u) - \langle p(t), f(t,x,u) \rangle + \langle \mu(t), \varphi(t,x,u) \rangle - \langle \dot{p}(t), x \rangle.$$

Observe that

$$F(t,x,u) = -H(t,x,u,p(t),\mu(t),1) - \langle \dot{p}(t), x \rangle$$

and, if $(x,u) \in Z_e(\mathcal{A}, f)$, then $K(x,u) = I(x,u)$. Let $(y,v) \in Y_1(x_0,u_0)$. By Note 2.3 we have

$$U_x(\tilde{x}_0(t))y(t) + U_u(\tilde{x}_0(t))v(t) = v(t) - \varphi_u^*(\tilde{x}_0(t))D(t)[\varphi_x(\tilde{x}_0(t))y(t) + \varphi_u(\tilde{x}_0(t))v(t)]$$

where $D(t) = [\varphi_u(\tilde{x}_0(t))\varphi_u^*(\tilde{x}_0(t))]^{-1}$. But the right-hand expression equals $v(t)$ and, therefore,

$$\begin{aligned} \dot{y}(t) &= A(t)y(t) + B(t)v(t) \\ &= \hat{A}(t)y(t) + \hat{B}(t)v(t). \end{aligned}$$

We conclude that $(y,v) \in Y_0(x_0,u_0)$ with respect to $\mathrm{P}(\mathcal{B}, \hat{f}, \hat{I})$. Also, by Lemma 3.6, (x_0,u_0) is normal to $\mathrm{P}(\mathcal{B}, \hat{f}, \hat{I})$. Therefore, by Lemma 1.4, there exist $\delta > 0$ and a one-parameter family $(\hat{x}(\cdot,\epsilon), \hat{u}(\cdot,\epsilon)) \in Z_e(\mathcal{B}, \hat{f})$ $(|\epsilon| < \delta)$ such that

i. $\hat{x}(t,0) = x_0(t)$, $\hat{u}(t,0) = u_0(t)$ $(t \in T)$.
ii. $\hat{x}_\epsilon(t,0) = y(t)$, $\hat{u}_\epsilon(t,0) = v(t)$ $(t \in T)$.

Set $x(t,\epsilon) := \hat{x}(t,\epsilon)$, $u(t,\epsilon) := U(t, \hat{x}(t,\epsilon), \hat{u}(t,\epsilon))$. Clearly $(x(\cdot,\epsilon), u(\cdot,\epsilon)) \in Z_e(\mathcal{A}, f)$. Let

$$g(\epsilon) := K(x(\cdot,\epsilon), u(\cdot,\epsilon)) = I(x(\cdot,\epsilon), u(\cdot,\epsilon)).$$

Thus $g(\epsilon) \geq g(0)$ $(|\epsilon| < \delta)$. Since

$$u_\epsilon(t,\epsilon) = U_x(t, \hat{x}(t,\epsilon), \hat{u}(t,\epsilon))\hat{x}_\epsilon(t,\epsilon) + U_u(t, \hat{x}(t,\epsilon), \hat{u}(t,\epsilon))\hat{u}_\epsilon(t,\epsilon)$$

we have

$$\begin{aligned} u_\epsilon(t,0) &= U_x(\tilde{x}_0(t))y(t) + U_u(\tilde{x}_0(t))v(t) \\ &= v(t) \end{aligned}$$

and, therefore,

$$\begin{aligned} 0 &\leq g''(0) = K''((x_0,u_0);(y,v)) \\ &= J_1((x_0,u_0);(y,v)). \ \blacksquare \end{aligned}$$

References:

[1] EG Gilbert, DS Bernstein, Second-order necessary conditions in optimal control: accessory-problem results without normality conditions, *Journal of Optimization Theory and Applications*, Vol 41, No 1, 1983, pp 75-106

[2] MR Hestenes, Calculus of Variations and Optimal Control Theory, John Wiley, New York, 1966

[3] J Warga, A second-order Lagrangian condition for restricted control problems, *Journal of Optimization Theory and Applications*, Vol 24, No 3, 1978, pp 475-483