Augmented Integrals for Optimal Control Problems

Javier F. Rosenblueth*

Abstract—By adding penalty functions to constrained minimum problems in finite dimensional spaces, one deals with unconstrained augmented problems for which the derivation of necessary optimality conditions can be obtained. Also, this technique yields in a natural way a method of multipliers for finding numerical solutions. In this paper we show how certain classes of optimal control problems with equality and inequality constraints can be treated in a similar way. A new notion of augmentability in optimal control is introduced and, without the usual assumption of normality, we derive first and second order necessary conditions for optimality.

Keywords—Augmentability, optimal control problems, equality and/or inequality constraints, normality

I. Introduction

In certain areas of optimization theory, the role of the theory of augmentability and penalty functions is well established. We refer to [10, 11] for an explanation of its importance in the literature. In particular, one type of augmentability can be seen as an alternative approach to that of regularity in the study of minimum problems involving equality and inequality constraints in finite dimensional spaces. For that kind of problems, a regularity assumption is usually imposed in order to derive the first and second order Lagrange multiplier rule (see [10]). However, it is generally rather difficult to verify if a certain point satisfies the notion of regularity and therefore one has to assume other simpler criteria to verify, such as that of normality, that imply regularity.

In the study of that kind of constrained minimum problems it is well known (see [10, 11]) that it is much simpler to derive both the first and second order Lagrange multiplier rules assuming that the problem is augmentable at a certain point than assuming instead either normality or regularity of the point. Another advantage of this approach is that it provides a method of multipliers used to find numerical solutions of constrained minimum problems. This method has been successfully generalized to a convex programming setting in [14]. The significance of this theory both in the finite dimensional case and in convex programming is well established (see, for example, [1, 2, 7, 8, 13–15, 20] and references therein, where a wide range of applications illustrate the use of the theory). However, this theory has received little attention in the development of other areas of optimization.

In Hestenes [10, 11], Rupp [19] and, more recently, in [16–18], several attempts to call attention to the role of augmentability in optimization theory have been made. In particular, in [17], a notion of augmentability was proposed for optimal control problems involving equality constraints both in the control and the state functions.

In this paper we shall generalize that notion for optimal control problems involving mixed equality and inequality constraints. Some fundamental properties of this kind of problems have been studied and we refer to [5] (and references therein) for a very general development of first order necessary conditions. The new notion of augmentability proposed in this paper provides an alternative approach to the development of not only first but also second order necessary conditions.

In order to clearly understand the type of augmentability we are dealing with, we shall first state some of the main aspects of the theories of regularity and augmentability for the finite dimensional case when equality and inequality constraints are present. We shall then summarize the main results applicable to optimal control problems with equality constraints and, finally, introduce the new notion for problems involving also inequality constraints.

*IIMAS–UNAM, Universidad Nacional Autónoma de México, Apartado Postal 20-726, México DF 01000, México (e-mail: jfrl@unam.mx).

II. THE FINITE DIMENSIONAL CASE

In this section we shall briefly explain the alternative approaches of regularity and augmentability which, for minimum problems with equality and inequality constraints in finite dimensional spaces, can be used in order to derive the first and second order Lagrange multiplier rule. We refer to Hestenes [10, 11] for a full account of the theory to follow.

A. Equality Constraints

Our starting point will be the case of minimum problems involving only equality constraints. As mentioned in the introduction, the augmentability approach is based on the removal of constraints. Thus we shall invoke well known optimality conditions for unconstrained problems, which are summarized in the following result.

Theorem 2.1 *Suppose we are given $S \subset \mathbf{R}^n$ open and a function $f\colon S \to \mathbf{R}$ of class C^2 on S. If x_0 affords a local minimum to f on S then $f'(x_0) = 0$, $f''(x_0) \geq 0$. Conversely, if $x_0 \in S$, $f'(x_0) = 0$ and $f''(x_0) > 0$, then there exist $\delta, m > 0$ such that*

$$|x - x_0| < \delta \Rightarrow f(x) \geq f(x_0) + m|x - x_0|^2.$$

Let us now sate the constrained problem we shall first study. Denote by $A = \{1, \ldots, m\}$ with $m < n$ and suppose we are given functions $f, g_\alpha\colon \mathbf{R}^n \to \mathbf{R}$ $(\alpha \in A)$ and

$$S = \{x \in \mathbf{R}^n \mid g_\alpha(x) = 0 \ (\alpha \in A)\}.$$

Let us consider the problem, which we label (P), of minimizing f on S.

For simplicity of exposition we shall assume that the functions f, $g = (g_1, \ldots, g_m)$ are of class C^2 on S but, as explained in [10], this assumption can be weakened.

For all $x_0 \in S$ define the set of *tangential constraints* by

$$R_S(x_0) := \{h \in \mathbf{R}^n \mid g'_\alpha(x_0; h) = 0 \ (\alpha \in A)\}.$$

A point x_0 is said to satisfy the *Lagrange multiplier rule* if it belongs to

$$\mathcal{L}(\lambda) := \{x_0 \in S \mid F'(x_0) = 0 \text{ and } F''(x_0; h) \geq 0 \text{ for all } h \in R_S(x_0)\}$$

where

$$F(x) := f(x) + \langle \lambda, g(x) \rangle$$

denotes the standard Lagrangian for problem (P).

Regularity

For all $x_0 \in S$ define the set $C_S(x_0)$ of *curvilinear tangent vectors of S at x_0* as the set of all $h \in \mathbf{R}^n$ for which there exist $\epsilon > 0$ and $x\colon (-\epsilon, \epsilon) \to S$ such that $x(0) = x_0$ and $\dot{x}(0) = h$.

As one readily verifies, $C_S(x_0) \subset R_S(x_0)$ for all $x_0 \in S$, but the converse may not hold.

Definition. A point $x_0 \in S$ is called a *regular point of S* if $C_S(x_0) = R_S(x_0)$.

First and second order necessary optimality conditions for the problem posed above, under the usual assumption of regularity, are the content of the following result. If that assumption is not imposed, one can easily find examples for which the optimality conditions on a local minimum to f on S may not hold.

Theorem 2.2 *If x_0 affords a local minimum to f on S and x_0 is a regular point of S then there exists $\lambda \in \mathbf{R}^m$ such that $x_0 \in \mathcal{L}(\lambda)$.*

This result requires the regularity assumption in order to assure that a local solution to the problem satisfies the Lagrange multiplier rule. However, it is in general difficult to test for regularity, and one simple widely used criterion is that of normality. A point $x_0 \in S$ is said to be a *normal point of S* if the linear equations

$$g'_\alpha(x_0; h) = 0 \quad (\alpha \in A)$$

in h are linearly independent. It is well known that normality implies regularity and, moreover, the multiplier λ in Theorem 2.2 is unique. It should be noted that the converse, however, may not hold so that the normality assumption is strictly stronger than that of regularity.

Augmentability

For any $(\lambda, \sigma) \in \mathbf{R}^m \times \mathbf{R}$, consider the function

$$H(x) = f(x) + \langle \lambda, g(x) \rangle + \sigma G(x)$$

where

$$G(x) = \frac{1}{2} \sum_1^m g_\alpha(x)^2$$

and define a set $\mathcal{A}(\lambda, \sigma)$ as the set of all $x_0 \in S$ such that x_0 affords a local minimum to H. Note that no constraints are present in this augmented problem.

Definition. We shall say that the problem (P) is *augmentable at x_0* if $x_0 \in \mathcal{A}(\lambda, \sigma)$ for some $(\lambda, \sigma) \in \mathbf{R}^m \times \mathbf{R}$.

Observe that $H(x) = f(x)$ for all $x \in S$. Clearly this implies that, if the problem (P) is augmentable at x_0, then x_0 affords a local minimum to f on S.

A derivation of the Lagrange multiplier rule can be obtained through the notion of augmentability. As the next results shows, if the problem (P) is augmentable at a point x_0, then the Lagrange multiplier rule holds at that point.

Theorem 2.3 *For all* $(\lambda, \sigma) \in \mathbf{R}^m \times \mathbf{R}$,

$$\mathcal{A}(\lambda, \sigma) \subset \mathcal{L}(\lambda).$$

Proof: Let $x_0 \in \mathcal{A}(\lambda, \sigma)$ and note that the following two equalities hold:

$$H(x) = F(x) + \sigma G(x),$$
$$S = \{x \in \mathbf{R}^n \mid G(x) = 0\}.$$

Since x_0 affords an unconstrained local minimum to H, it follows by Theorem 2.1 that

$$H'(x_0) = 0 \text{ and } H''(x_0) \geq 0.$$

Since $G(x_0) = 0$ and $G(x) \geq 0$ for all $x \in \mathbf{R}^n$, x_0 minimizes G. We invoke Theorem 2.1 once more to conclude that $G'(x_0) = 0$. Therefore,

$$0 = H'(x_0) = F'(x_0) + \sigma G'(x_0) = F'(x_0),$$
$$0 \leq H''(x_0) = F''(x_0) + \sigma G''(x_0)$$

and so $F''(x_0; h) \geq 0$ whenever $G''(x_0; h) = 0$.

Now, since

$$G''(x_0; h) = \sum_1^m g'_\alpha(x_0; h)^2,$$

we have $F''(x_0; h) \geq 0$ whenever

$$g'_\alpha(x_0; h) = 0 \quad (\alpha \in A),$$

and so $x_0 \in \mathcal{L}(\lambda)$. ∎

Though we shall not treat sufficiency in this paper, it is important to mention that the classical sufficient conditions for optimality imply augmentability. Since a point x_0 affords a local minimum to f on S if the problem (P) is augmentable at x_0, this result provides an alternative way of proving the classical sufficiency theorem. It also justifies the use of the augmentability approach in the theory of optimality conditions.

A point x_0 is said to satisfy the *strengthened Lagrange multiplier rule* if it belongs to

$$\mathcal{L}'(\lambda) := \{x_0 \in S \mid F'(x_0) = 0 \text{ and } F''(x_0; h) > 0 \text{ for all } h \in R_S(x_0),\ h \neq 0\}$$

and, as one can easily verify, the following result holds (see [10]) by a simple application of Theorem 2.1.

Theorem 2.4 *Suppose* $x_0 \in \mathcal{L}'(\lambda)$ *for some* $\lambda \in \mathbf{R}^m$. *Then there exist* σ_0, $k > 0$ *and a neighborhood* N *of* x_0 *such that, for all* $\sigma \geq \sigma_0$ *and* $x \in N$,

$$H(x) \geq H(x_0) + k|x - x_0|^2.$$

In particular, (P) *is augmentable at* x_0 *and, for all* $x \in S \cap N$,

$$f(x) \geq f(x_0) + k|x - x_0|^2.$$

To illustrate these two approaches, consider the problem of minimizing $f\colon \mathbf{R}^2 \to \mathbf{R}$ on the set

$$S = \{(x, y) \in \mathbf{R}^2 \mid g(x, y) = 0\}.$$

Example 2.5 *The point* $(0, 0)$ *is not regular with respect to problem* (P) *but the problem is augmentable at this point.*

Consider the functions

$$f(x, y) = x^2 - y^4 - 4y^2, \quad g(x, y) = y^2.$$

Clearly $(0, 0)$ is not a regular point of S since $C_S((0, 0))$ coincides with the line $y = 0$, while $R_S((0, 0))$ is $\mathbf{R}^2$.

However, the problem is augmentable at $(0, 0)$ since

$$H(x, y) = x^2 + \left(\frac{\sigma - 2}{2}\right)y^4 + (\lambda - 4)y^2$$

is minimized at the origin with $\lambda = 4$ and $\sigma > 2$.

Example 2.6 *The point* $(0, 0)$ *is regular with respect to problem* (P) *but the problem is not augmentable at this point.*

Consider the functions

$$f(x, y) = x^2 + 2x + y^4, \quad g(x, y) = xy - x.$$

It is clear that f has a strict local minimum at $(0, 0)$ on S. Since $g'(0, 0) \neq (0, 0)$, it is a normal point and hence regular for (P).

However, the problem is not augmentable at $(0, 0)$. To prove it, note that

$$H(x, y) = x^2\left(1 + \frac{\sigma}{2}(y - 1)^2\right) + (2 - \lambda)x + \lambda xy + y^4.$$

If $H'(0, 0) = (0, 0)$, we must have $\lambda = 2$. Hence,

$$H(x, y) = x^2\left(1 + \frac{\sigma}{2}(y - 1)^2\right) + 2xy + y^4.$$

We can suppose that $\sigma \geq 0$. If $(y-1)^2 \leq 4$, $y^2 < 1/(1+2\sigma)$ and $x = -y/(1+2\sigma)$, we see that

$$H(x,y) \leq y^2\left(y^2 - \frac{1}{1+2\sigma}\right) < 0 = H(0,0)$$

and so (P) is not augmentable at $(0,0)$.

Let us briefly mention that one of the main advantages of the augmentability approach is that it provides in a natural way a method for solving the original problem. Using the notation

$$H(x,\lambda,\sigma) = f(x) + \langle \lambda, g(x)\rangle + \frac{\sigma}{2}|g(x)|^2$$

the method consists in choosing λ_0 and $\sigma > 0$, hopefully so that $H(x,\lambda_0,\sigma)$ is convex in x. Select $\xi_0, \xi_1, \ldots$ with $\xi_k \geq \xi_0 > 0$ and choose x_k, λ_k successively so that x_k minimizes

$$H(x, \lambda_{k-1}, \sigma + \xi_{k-1})$$

and set

$$\lambda_k = \lambda_{k-1} + \xi_{k-1} g(x_k).$$

Then, as explained in [11], usually $\{x_k\}$ converges to a solution x_0 of the original problem. Moreover, in this case, $\{\lambda_k\}$ converges to the Lagrange multiplier associated with x_0.

B. Inequality Constraints

Let us now present the main features related to the regularity and augmentability approaches for problems involving equality and inequality constraints. This will allow us to compare it with the notion of augmentability introduced in the following section for optimal control problems.

Suppose we are given functions f mapping $\mathbf{R}^n$ to $\mathbf{R}$ and $g = (g_1, \ldots, g_m)$ mapping $\mathbf{R}^n$ to $\mathbf{R}^m$. Let now

$$S := \{x \in \mathbf{R}^n \mid g_\alpha(x) \leq 0 \ (\alpha \in A), \quad g_\beta(x) = 0 \ (\beta \in B)\}$$

where $A = \{1, \ldots, p\}$ and $B = \{p+1, \ldots, m\}$, and consider the problem, which we label (Q), of minimizing f on S. As before, we shall assume that f and g are $C^2(S)$.

For this problem, the *Lagrangian* (with respect to λ in $\mathbf{R}^m$) coincides with the one for equality constraints, that is,

$$F(x) = f(x) + \langle \lambda, g(x)\rangle \quad (x \in \mathbf{R}^n).$$

We define the set of *Lagrange multipliers* at $x \in S$ as

$$P(x) = \{\lambda \in \mathbf{R}^m \mid \lambda_\alpha \geq 0 \ (\alpha \in A), \quad \lambda_\alpha = 0 \text{ if } g_\alpha(x) < 0\}.$$

The set $R_S(x)$ of *tangential constraints* at $x \in S$ corresponds to

$$\{h \in \mathbf{R}^n \mid g'_\alpha(x;h) \leq 0 \ (\alpha \in A,\ g_\alpha(x) = 0), \quad g'_\beta(x;h) = 0 \ (\beta \in B)\},$$

and the set $\tilde{R}_S(x,\lambda)$ of *modified tangential constraints* is given by

$$\{h \in R_S(x) \mid g'_\alpha(x;h) = 0 \ (\alpha \in A,\ \lambda_\alpha > 0)\}.$$

A point x_0 is said to satisfy the *Lagrange multiplier rule* if it belongs to

$$\mathcal{L}(\lambda) := \{x_0 \in S \mid \lambda \in P(x_0),\ F'(x_0) = 0 \text{ and } F''(x_0;h) \geq 0 \text{ for all } h \in \tilde{R}_S(x_0,\lambda)\}$$

Note that the second order condition holds not on the set of tangential constraints but on the modified one.

Regularity

For this problem we begin by recalling the notion of tangent cone of a set in $\mathbf{R}^n$ at a given point.

A sequence $\{x_q\} \subset \mathbf{R}^n$ will be said to *converge to* x_0 *in the direction* h if h is a unit vector, $x_q \neq x_0$, and

$$\lim_{q\to\infty} |x_q - x_0| = 0, \quad \lim_{q\to\infty} \frac{x_q - x_0}{|x_q - x_0|} = h.$$

Given $x_0 \in C \subset \mathbf{R}^n$, the *tangent cone of* C *at* x_0, denoted by $T_C(x_0)$, is the (closed) cone determined by the unit vectors h for which there exists a sequence $\{x_q\}$ in C converging to x_0 in the direction h.

One can find other equivalent definitions of this set but the one we are using corresponds to that of Hestenes [9, 10].

Now, as one readily verifies, for the set S of constraints delimiting our problem (Q), $T_S(x_0) \subset R_S(x_0)$ for all $x_0 \in S$ but the converse may not hold.

Definition. A point $x_0 \in S$ will be called a *regular point of* S if $T_S(x_0) = R_S(x_0)$.

The regularity approach, as in the case of the previous problem (P), provides first order optimality conditions for problem (Q), and it can be stated as follows.

Theorem 2.7 *If x_0 affords a local minimum to f on S and x_0 is a regular point of S then there exists $\lambda \in P(x_0)$ such that $F'(x_0) = 0$.*

Second order necessary conditions hold again if a regularity assumption is imposed on a local solution satisfying the first order Lagrange multiplier. This assumption, however, is not imposed on the set S but on a subset of S which takes into account positive Lagrange multipliers. The following result summarizes those conditions.

Theorem 2.8 *Let $x_0 \in S$ and suppose there exists $\lambda \in P(x_0)$ such that $F'(x_0) = 0$. Let*

$$\tilde{S} := \{x \in S \mid g_\alpha(x) = 0 \ (\alpha \in A, \ \lambda_\alpha > 0)\}.$$

If x_0 affords a local minimum to f on S and x_0 is a regular point of $\tilde{S}$ then $x_0 \in \mathcal{L}(\lambda)$.

The derivation of the Lagrange multiplier rule through this approach usually requires several fundamental results on the theory of convex cones. On the other hand, it is generally difficult to test for regularity and, as in the case of equality constraints, one criterion is that of normality. A point $x_0 \in S$ is said to be a *normal point of* S if the relations

$$\sum_{i=1}^{m} \lambda_i g_i'(x_0) = 0 \quad \text{and} \quad \lambda \in P(x_0)$$

imply that $\lambda = 0$. One can then prove, by making use of the implicit function theorem, that normality implies regularity.

Augmentability

Let us turn now to the augmentability approach. As explained in the introduction, the main purpose of this approach is to simplify the original problem by removing constraints.

To begin with, let

$$\hat{S} = \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^p \mid y^\alpha \le 0, \\ g_\alpha(x) - y^\alpha = 0 \ (\alpha \in A), \ g_\beta(x) = 0 \ (\beta \in B)\}$$

and, for all $(\lambda, \sigma) \in \mathbf{R}^m \times \mathbf{R}$, define

$$H(x, y) = f(x) + \sum_1^p \lambda_\alpha (g_\alpha(x) - y^\alpha) + \sum_{p+1}^m \lambda_\beta g_\beta(x) + \sigma G(x, y)$$

where

$$G(x, y) = \frac{1}{2}\Big(\sum_1^p \{g_\alpha(x) - y^\alpha\}^2 + \sum_{p+1}^m g_\beta(x)^2\Big).$$

Note that we can express $\hat{S}$ as

$$\hat{S} = \{(x, y) : G(x, y) = 0, \ y^\alpha \le 0 \ (\alpha \in A)\}.$$

Observe that

$$H(x, y) = f(x)$$

whenever $(x, y) \in \hat{S}$. Consequently, if $x_0 \in S$ and $y_0^\alpha = g_\alpha(x_0)$ $(\alpha \in A)$, (x_0, y_0) affords a local minimum to H on $\hat{S}$ if and only if x_0 is a local solution to our original problem (Q).

Definition. We shall say that (Q) is *augmentable* at $x_0 \in S$ if there exists $(\lambda, \sigma) \in \mathbf{R}^m \times \mathbf{R}$ such that (x_0, y_0), with $y_0^\alpha = g_\alpha(x_0)$ $(\alpha \in A)$, affords a local minimum to H on the set

$$K = \{(x, y) \in \mathbf{R}^n \times \mathbf{R}^p \mid y^\alpha \le 0 \ (\alpha \in A)\}.$$

Clearly, if (Q) is augmentable at x_0, then x_0 affords a local minimum to f on S.

Let us state necessary conditions for the problem of minimizing H on K. Define the augmented Lagrangian with respect to f, g and $\lambda \in \mathbf{R}^m$ by

$$\hat{F}(x, y) = f(x) + \langle \lambda, g(x) \rangle - \sum_1^p \lambda_\alpha y^\alpha$$

and observe that

$$H(x, y) = \hat{F}(x, y) + \sigma G(x, y).$$

Let $C_{\hat{F}}(x, y)$ to be the set of points (h, k) in $\mathbf{R}^n \times \mathbf{R}^p$ such that

$$k^\alpha \le 0 \text{ if } y^\alpha = 0,$$

and

$$k^\alpha = 0 \text{ if } \hat{F}_{y^\alpha}(x, y) = -\lambda_\alpha < 0.$$

Theorem 2.9 *Let $(x_0, y_0) \in \hat{S}$ and suppose that, for some $\sigma \in \mathbf{R}$, (x_0, y_0) affords a local minimum to H on K. Then (x_0, y_0) affords a local minimum to $\hat{F}$ on $\hat{S}$ and the following holds:*

a. $\hat{F}_x(x_0, y_0) = 0$, $\hat{F}_{y^\alpha}(x_0, y_0) \le 0$ *with equality if* $y_0^\alpha < 0$.

b. $\hat{F}''((x_0, y_0); (h, k)) \ge 0$ *for all* (h, k) *in* $C_{\hat{F}}(x_0, y_0)$ *satisfying* $G''((x_0, y_0); (h, k)) = 0$.

Based on this result, one can easily prove (see [10]) that the Lagrange multiplier rule is a consequence of augmentability.

Theorem 2.10 *Let $x_0 \in S$ and suppose* (Q) *is augmentable at x_0. Then the Lagrange multiplier rule* *holds at* x_0. *Moreover,* x_0 *affords a local minimum to* f *on* S.

As in the case of only equality constraints, the classical sufficient conditions imply augmentability. A point is said to satisfy the *strengthened Lagrange multiplier rule* if it belongs to

$$\mathcal{L}'(\lambda) := \{x_0 \in S \mid \lambda \in P(x_0),\ F'(x_0) = 0 \text{ and } F''(x_0; h) > 0 \text{ for all } h \neq 0 \text{ in } \in \tilde{R}_S(x_0, \lambda)\}$$

Theorem 2.11 *Let* $x_0 \in S$ *and suppose that, with respect to* λ, *the strengthened Lagrange multiplier rule holds at* x_0. *Set* $y_0^\alpha = g_\alpha(x_0)$. *Then there exist* σ_0, $k > 0$ *and a neighborhood* N *of* x_0 *such that, if* $\sigma \geq \sigma_0$, $x \in N$ *and* $(x, y) \in K$, *then*

$$H(x, y) \geq H(x_0, y_0) + k|x - x_0|^2.$$

In particular, (Q) *is augmentable at* x_0 *and, for all* $x \in N \cap S$,

$$f(x) \geq f(x_0) + k|x - x_0|^2.$$

III. Optimal Control Problems

The previous theory may provide some ideas in trying to remove constraints for other optimization problems. In this section we shall give a possible direction in that respect for certain classes of optimal control problems. The main novelty of this paper corresponds to problems involving equality and inequality constraints.

It is worth mentioning that the role of penalty functions in optimal control has been used to find solutions to the problem and in the derivation of necessary conditions (see [3] for a detailed explanation). To illustrate the technique used in [3], consider an optimal control problem where the cost is given by

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

and constraints in the state are given by inequalities of the type

$$h(t, x(t)) \leq 0 \quad \text{a.e. in } [t_0, t_1].$$

Then the constraints are removed by penalizing the cost with the integral

$$\int_{t_0}^{t_1} \max\{0, h(t, x(t))\}dt$$

thus obtaining a sequence of problems where one is interested in minimizing

$$\int_{t_0}^{t_1} L(t, x(t), u(t))dt + K \int_{t_0}^{t_1} \max\{0, h(t, x(t))\}dt$$

without constraints in the state functions.

This technique produces a nonsmooth optimal control problem since the the cost with the penalty term is not differentiable. For that kind of problems, first order optimality conditions are well established (see, for example, [5]). However, in this paper, we shall propose an augmentable integral with a penalty term which is differentiable and for which first and also second order conditions are obtainable.

We also refer to [4] for a non-variational method for solving a linear quadratic optimal control problem involving inequality constraints which can be treated by using the results of this section.

A. Equality Constraints

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

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

where φ is a function mapping $T \times \mathbf{R}^n \times \mathbf{R}^m$ to $\mathbf{R}^q$ ($q \leq m$). We assume that φ is C^2 on $\mathcal{A}$ and the matrix $\varphi_u(t, x, u)$ has rank q on $\mathcal{A}$.

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 $Z := X \times \mathcal{U}$,

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

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

and

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

The problem we shall deal with, which we label (CP), is that of minimizing I over $Z_e(\mathcal{A})$.

The elements of Z are called *processes*, of $Z_e(\mathcal{A})$ *admissible processes*, and a process (x, u) *solves* (CP) if (x, u) is admissible and $I(x, u) \leq I(y, v)$ for all admissible processes (y, v).

Assume that f, L are C^2 on $\mathcal{A}$ and denote by '$*$' the transpose.

Denote by $\mathcal{U}_r$ the space of piecewise continuous functions mapping T to $\mathbf{R}^r$ ($r \in \mathbf{N}$).

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

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

For any $(x, u, p, \mu) \in Z \times X \times \mathcal{U}_q$ and $(y, v) \in Z$ let

$$J((x, u, p, \mu); (y, v)) := \int_{t_0}^{t_1} 2\Omega(t, y(t), v(t))dt$$

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

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

and $H(t)$ denotes $H(t, x(t), u(t), p(t), \mu(t))$.

Normality

Definition. A process (x, u) is said to be *normal* if, given $p \in X$ and $\mu \in \mathcal{U}_q$ satisfying

$$\dot{p}(t) = -A^*(t)p(t) + \varphi_x^*(t, x(t), u(t))\mu(t),$$

$$0 = B^*(t)p(t) - \varphi_u^*(t, x(t), u(t))\mu(t)$$

then $p \equiv 0$, where

$$A(t) := f_x(t, x(t), u(t)),\ B(t) := f_u(t, x(t), u(t)).$$

A derivation of the following set of necessary conditions, under normality assumptions, can be found in [6]. In the remaining of this paper we shall denote by $[t]$ the point $(t, x_0(t), u_0(t))$.

Theorem 3.1 *Suppose* (x_0, u_0) *is a normal solution of* (CP). *Then there exists* $(p, \mu) \in X \times \mathcal{U}_q$ *such that*

$$\dot{p}(t) = -H_x^*(t, x_0(t), u_0(t), p(t), \mu(t)),$$

$$H_u(t, x_0(t), u_0(t), p(t), \mu(t)) = 0 \quad (t \in T).$$

Moreover,

$$J((x_0, u_0, p, \mu); (y, v)) \geq 0$$

for all $(y, v) \in Z$ *satisfying*

$$y(t_0) = y(t_1) = 0,$$

$$\dot{y}(t) = f_x[t]y(t) + f_u[t]v(t) \quad (t \in T),$$

$$\varphi_x[t]y(t) + \varphi_u[t]v(t) = 0 \quad (t \in T).$$

Augmentability

Associated with the integral I, consider the augmented integral

$$K(x, u; \mu, \sigma) = \int_{t_0}^{t_1} F(t, x(t), u(t))dt$$

where

$$F(t, x, u) = L(t, x, u) + \langle \mu(t), \varphi(t, x, u)\rangle + \sigma(t, x, u)G(t, x, u),$$

$$G(t, x, u) = \frac{1}{2}\sum_1^q \varphi_i(t, x, u)^2.$$

Definition. We shall say that (CP) is *augmentable at* (x_0, u_0) if there exists (μ, σ) such that (x_0, u_0) solves the unconstrained problem of minimizing $K(x, u; \mu, \sigma)$ over all $(x, u) \in D$ with $x(t_0) = \xi_0$, $x(t_1) = \xi_1$, and is normal with respect to that problem, that is, $z \equiv 0$ is the only solution of the system

$$\dot{z}(t) = -A^*[t]z(t), \quad B^*[t]z(t) = 0.$$

Note that, in this event, (x_0, u_0) solves (CP) since, for any (x, u) admissible process for (CP), we have

$$I(x_0, u_0) = K(x_0, u_0) \leq K(x, u) = I(x, u).$$

With this definition of augmentability the following result holds (see [17]).

Theorem 3.2 *Suppose* (CP) *is augmentable at* (x_0, u_0) *with respect to* (μ, σ). *Then there exists* $p \in X$ *such that* (p, μ) *satisfies the conclusions of Theorem 3.1.*

The following simple example shows that a solution of a problem may not be normal but the problem is augmentable at that point.

Example 3.3 Consider the problem of minimizing

$$I(x, u) = \int_0^{\pi} \{u^2(t) - x^2(t)\}dt$$

over all $(x, u) \in X \times \mathcal{U}$ satisfying $\dot{x} = u$, $x(0) = x(\pi) = 0$ and $\sin u(t) = 0$ $(t \in T)$.

Note first that $(x_0, u_0) \equiv (0, 0)$ solves the problem. Now, from the definition of A and B we have $A(t) = 0$ and $B(t) = 1$. Clearly (x_0, u_0) is not normal since, if (p, μ) satisfies

$$\dot{p}(t) = -A(t)p(t) + \varphi_x[t]\mu(t) = 0,$$

$$0 = B(t)p(t) - \varphi_u[t]\mu(t) = p(t) - \mu(t)$$

then the condition $p \equiv 0$ does not necessarily follows. Thus we cannot apply Theorem 3.1.

On the other hand, the function F is given by

$$u^2 - x^2 + \mu(t)\,\sin u + \tfrac{1}{2}\sigma(t,x,u)\,\sin^2 u.$$

Therefore, if $\mu \equiv \sigma \equiv 0$, then (x_0, u_0) solves the unconstrained problem of minimizing

$$K(x,u;0,0) = \int_0^\pi \{u^2(t) - x^2(t)\}dt$$

over all $(x,u) \in X \times \mathcal{U}$ satisfying $\dot{x} = u$, $x(0) = x(\pi) = 0$. Also, $z \equiv 0$ is the only solution of the system

$$\dot{z}(t) = -A(t)z(t) = 0, \quad B(t)z(t) = z(t) = 0.$$

Hence the problem is augmentable at (x_0, u_0), and an application of Theorem 3.2 yields first and second order necessary conditions.

B. Inequality Constraints

Suppose the data are as before except for the set of constraints which is now given by

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

$$\varphi_\alpha(t,x,u) \le 0 \ (\alpha \in R),\ \varphi_\beta(t,x,u) = 0 \ (\beta \in Q)\}$$

where $R = \{1,\ldots,r\}$, $Q = \{r+1,\ldots,q\}$.

Assume that the function $\varphi\colon T \times \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}^q$ given by $\varphi = (\varphi_1,\ldots,\varphi_q)$ is of class C^2 and the $q \times (m+r)$-dimensional matrix

$$\left(\frac{\partial \varphi_i}{\partial u^k} \ \ \delta_{i\alpha}\varphi_\alpha\right)$$

$$(i = 1,\ldots,q;\ \alpha = 1,\ldots,r;\ k = 1,\ldots,m)$$

has rank q on $\mathcal{A}$.

This condition is equivalent to the condition that at each point (t,x,u) in $\mathcal{A}$, the matrix

$$\left(\frac{\partial \varphi_i}{\partial u^k}\right) \qquad (i = i_1,\ldots,i_p;\ k = 1,\ldots,m)$$

has rank p, where $i_1,\ldots,i_p$ are the indices $i \in \{1,\ldots,q\}$ such that $\varphi_i(t,x,u) = 0$.

Consider the mixed equality/inequality constrained optimal control problem, which we label (CQ), of minimizing I over $Z_e(\mathcal{A})$.

Normality

Denote by $\mathcal{V}(x,u)$ the set of all multipliers $\mu \in \mathcal{U}_q$ such that $\mu_\alpha(t) = 0$ whenever $\varphi_\alpha(t,x(t),u(t)) < 0$.

Definition. A process (x,u) will be said to be *normal* to (CQ) if, given $(p,\mu) \in X \times \mathcal{V}(x,u)$ such that, for all $t \in T$,

$$\dot{p}(t) = -A^*(t)p(t) + \varphi_x^*(t,x(t),u(t))\mu(t)$$

$$0 = B^*(t)p(t) - \varphi_u^*(t,x(t),u(t))\mu(t)$$

then $p \equiv 0$.

We refer to [6, 12] for the derivation of the following set of necessary conditions for problem (CQ). Let $\hat{\varphi} = (\varphi_{i_1},\ldots,\varphi_{i_p})$ where $i_1,\ldots,i_p$ are the active indices at $[t]$, that is, those $i \in R \cup Q$ such that $\varphi_i(t,x_0(t),u_0(t)) = 0$.

Theorem 3.4 *Suppose* (x_0,u_0) *is a normal solution of* (CQ). *Then there exists* $(p,\mu) \in X \times \mathcal{V}(x_0,u_0)$ *with* $\mu_\alpha \ge 0$ $(\alpha \in R)$ *such that*

$$\dot{p}(t) = -H_x^*(t,x_0(t),u_0(t),p(t),\mu(t)),$$

$$H_u(t,x_0(t),u_0(t),p(t),\mu(t)) = 0 \quad (t \in T).$$

Moreover,

$$J((x_0,u_0,p,\mu);(y,v)) \ge 0$$

for all $(y,v) \in Z$ *satisfying* $y(t_0) = y(t_1) = 0$,

$$\dot{y}(t) = f_x[t]y(t) + f_u[t]v(t) \quad (t \in T),$$

$$\hat{\varphi}_x[t]y(t) + \hat{\varphi}_u[t]v(t) = 0 \quad (t \in T).$$

Augmentability

Associated with the integral I, consider the augmented integral

$$K(x,u,b;\mu,\sigma) = \int_{t_0}^{t_1} F(t,x(t),u(t),b(t))dt$$

where

$$F(t,x,u,b) = L(t,x,u) + \langle \mu(t), \psi(t,x,u,b)\rangle$$

$$+\ \sigma(t,x,u,b)G(t,x,u,b),$$

$$G(t,x,u,b) = \frac{1}{2}\sum_1^q \psi_i(t,x,u,b)^2,$$

and the function ψ is given by

$$\psi_i(t,x,u,b) = \begin{cases} \varphi_i(t,x,u) - b^i & \text{if } i \in R \\ \varphi_i(t,x,u) & \text{if } i \in Q. \end{cases}$$

Consider now the problem (P_{aug}) of minimizing $K(x,u,b;\mu,\sigma)$ subject to

a. $(x,u,b) \in X \times \mathcal{U}_m \times \mathcal{U}_r$.
b. $\dot{x}(t) = \hat{f}(t,x(t),u(t),b(t))$ $(t \in T)$.
c. $x(t_0) = \xi_0$, $x(t_1) = \xi_1$.
d. $b^\alpha(t) \le 0$ $(\alpha \in R,\ t \in T)$

where $\hat{f}(t,x,u,b) = f(t,x,u)$.

Define $\mathcal{B}$ as the set of all (t,x,u,b) in $T \times \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^r$ satisfying $b^\alpha \le 0$ $(\alpha \in R)$ and $\psi(t,x,u,b) = 0$. Observe that

$$K(x,u,b;\mu,\sigma) = I(x,u)$$

whenever $(t,x(t),u(t),b(t)) \in \mathcal{B}$. Consequently, for any (μ,σ), if $(x_0,u_0) \in Z_e(\mathcal{A})$ and $b_0^\alpha = \varphi(t,x_0(t),u_0(t))$ $(\alpha \in R,\ t \in T)$, then the triple (x_0,u_0,b_0) solves the problem of minimizing K on the set of all $(x,u,b) \in X \times \mathcal{U}_m \times \mathcal{U}_r$ satisfying

$$\dot{x}(t) = \hat{f}(t,x(t),u(t),b(t)) \quad (t \in T),$$

$$x(t_0) = \xi_0, \quad x(t_1) = \xi_1,$$

$$(t,x(t),u(t),b(t)) \in \mathcal{B} \quad (t \in T)$$

if and only if (x_0,u_0) solves the original problem (CQ).

The definition of augmentability we propose, based on the one introduced for the finite dimensional case, is the following.

Definition. We shall say that (CQ) is *augmentable at* (x_0,u_0) if there exists (μ,σ) such that (x_0,u_0,b_0) with $b_0^\alpha(t) := \varphi_\alpha(t,x_0(t),u_0(t))$ solves the augmented problem (P_{aug}) and is normal with respect to that problem.

Clearly, if (CQ) is augmentable at (x_0,u_0), then (x_0,u_0) is a solution to the problem (CQ).

Based on this definition, we obtain the following result.

Theorem 3.5 *Let* $(x_0,u_0) \in Z_e(\mathcal{A})$ *and suppose* (CQ) *is augmentable at* (x_0,u_0) *with respect to* (μ,σ)*. Then there exists* $p \in X$ *such that* (p,μ) *satisfies the conclusions of Theorem 3.4.*

Proof: Let us begin by defining the Hamiltonian corresponding to the augmented problem, that is,

$$\tilde{H}(t,x,u,b,p,\nu) =$$

$$\langle p, \hat{f}(t,x,u,b)\rangle - F(t,x,u,b) - \langle \nu, \beta\rangle$$

where $\beta_\alpha(t,x,u,b) = b^\alpha$ $(\alpha \in R)$. Denote by $\tilde{\mathcal{V}}(b)$ the set of multipliers $\nu \in \mathcal{U}_r$ such that $\nu_\alpha(t) = 0$ whenever $b^\alpha(t) < 0$.

Since (CQ) is augmentable at (x_0,u_0) with respect to (μ,σ), (x_0,u_0,b_0) with $b_0^\alpha(t) := \varphi_\alpha(t,x_0(t),u_0(t))$ solves the augmented problem (P_{aug}) and is normal with respect to that problem. An application of Theorem 3.4 yields the existence of $(p,\nu) \in X \times \tilde{\mathcal{V}}(b_0)$ with $\nu_\alpha \ge 0$ $(\alpha \in R)$ such that

$$\dot{p}(t) = -\tilde{H}_x^*(t,x_0(t),u_0(t),b_0(t),p(t),\nu(t)),$$

$$\tilde{H}_u(t,x_0(t),u_0(t),b_0(t),p(t),\nu(t)) = 0 \quad (t \in T)$$

and, moreover,

$$\tilde{J}((x_0,u_0,b_0,p,\nu);(y,v)) \ge 0$$

for all $(y,v) \in Z$ satisfying $y(t_0) = y(t_1) = 0$,

$$\dot{y}(t) = \hat{f}_x[t]y(t) + \hat{f}_{(u,b)}[t]v(t) \quad (t \in T),$$

$$\beta_x[t]y(t) + \beta_{(u,b)}[t]v(t) = 0 \quad (t \in T)$$

where we have used the notation

$$\tilde{J}((x,u,b,p,\nu);(y,v)) = \int_{t_0}^{t_1} 2\tilde{\Omega}(t,y(t),v(t))dt,$$

$2\tilde{\Omega}(t,y,v)$ is given by

$$-\,[\langle y, \tilde{H}_{xx}(t)y\rangle + \ 2\langle y, \tilde{H}_{xub}(t)v\rangle + \langle v, \tilde{H}_{ubub}(t)v\rangle]$$

and $\tilde{H}(t)$ denotes $\tilde{H}(t,x(t),u(t),b(t),p(t),\nu(t))$.

Since $\psi(t,x_0(t),u_0(t),b_0(t)) = 0$ for all $t \in T$, one readily verifies that

$$\tilde{H}_x(t) = H_x(t) \quad \text{and} \quad \tilde{H}_u(t) = H_u(t),$$

and so

$$\dot{p}(t) = -H_x^*(t,x_0(t),u_0(t),p(t),\mu(t)),$$

$$H_u(t,x_0(t),u_0(t),p(t),\mu(t)) = 0 \quad (t \in T).$$

On the other hand, we have

$$\tilde{J}((x_0,u_0,b_0(t),p,\nu);(y,v)) =$$

$$J((x_0,u_0,p,\mu);(y,v)) + \int_{t_0}^{t_1} \sigma(t,x_0(t),u_0(t),b_0(t))$$

$$|\varphi_x(t,x_0(t),u_0(t))y(t) + \varphi_u(t,x_0(t),u_0(t))v(t)|^2 dt$$

and therefore

$$J((x_0,u_0,p,\mu);(y,v)) \ge 0$$

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

$$y(t_0) = y(t_1) = 0,$$

$$\dot{y}(t) = f_x[t]y(t) + f_u[t]v(t) \quad (t \in T),$$

$$\hat{\varphi}_x[t]y(t) + \hat{\varphi}_u[t]v(t) = 0 \quad (t \in T). \blacksquare$$

IV. Conclusions

This paper provides a new approach to the derivation of first and second order necessary conditions for certain constrained optimization problems. It is based on the theory of augmentability which has been successfully applied to finite dimensional problems and convex programming. In the former, it is well-known that it is much simpler to derive the Lagrange multiplier rule under an assumption of augmentability than under the assumption of regularity. With the idea of illustrating this approach for problems involving equality and inequality constraints, a brief summary of the main techniques and results is given.

For optimal control problems involving mixed equality and inequality constraints, a notion of augmentability is proposed and optimality conditions are derived. It is of interest to see if the set of variations where the second order conditions hold can be enlarged through this approach, to obtain conditions which imply augmentability, and to derive a method of multipliers for finding numerical solutions of such problems.

References

[1] D.P. Bertsekas, On penalty and multiplier methods for constrained minimization, *SIAM Journal on Control and Optimization* 14, 1976, pp. 216-235.

[2] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press, New York 1982.

[3] H.A. Biswas, Necessary conditions for optimal control problems with and without state constraints: a comparative study, *WSEAS Transactions on Systems and Control* 6, 2011, pp. 217-228.

[4] C. Botan and F. Ostafi, A solution to the continuous and discrete-time linear quadratic optimal problems, *WSEAS Transactions on Systems and Control* 3, 2008, pp. 71-78.

[5] F. Clarke and M.R. de Pinho, Optimal control problems with mixed constraints, *SIAM Journal on Control and Optimization* 48, 2010, pp. 4500-4524.

[6] M.R. de Pinho and J.F. Rosenblueth, Mixed constraints in optimal control: an implicit function theorem approach, *IMA Journal of Mathematical Control and Information* 24, 2007, pp. 197–218.

[7] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods. Applications to the Numerical Solution of Boundary-Value Problems*, North-Holland, Amsterdam 1983.

[8] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer Series in Computational Physics, Springer-Verlag, New York 1984.

[9] M.R. Hestenes, *Calculus of Variations and Optimal Control Theory*, John Wiley & Sons, New York 1966.

[10] M.R. Hestenes, *Optimization Theory, The Finite Dimensional Case*, John Wiley & Sons, New York 1975.

[11] M.R. Hestenes, Augmentability in optimization theory, *Journal of Optimization Theory and Applications* 32, 1980, pp. 427–440.

[12] P.D. Loewen and H. Zheng, Generalized conjugate points for optimal control problems, *Nonlinear Analysis, Theory, Methods & Applications* 22, 1994, pp. 771–791.

[13] A. Miele, P.E. Moseley, A.V. Levy and G.M. Coggins, On the method of multipliers for mathematical programming problems, *Journal of Optimization Theory and Applications* 10, 1972, pp. 1-33.

[14] R.T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, *Journal of Optimization Theory and Applications* 12, 1973, pp. 555–562.

[15] R.N. Rodrigues, E.L. da Silva, E.C. Finardi and F. Takigawa, Solving the short-term scheduling problem of hydrothermal systems via Lagrangian relaxation and augmented Lagrangian, *Mathematical Problems in Engineering*, 2012, Art. ID: 856178.

[16] J.F. Rosenblueth, Admissible variations for optimal control problems with mixed constraints, *WSEAS Transactions on Systems* 4, 2005, pp. 2204–2211.

[17] J.F. Rosenblueth, Augmentability in optimal control, *International Journal of Mathematics and Statistics* 5, 2009, pp. 102–109.

[18] J.F. Rosenblueth, Constrained optimization through augmentability, *10th WSEAS International Conference on Dynamical Systems and Control*, Iasi, Romania, WSEAS Press 2011, pp. 105-112.

[19] R.D. Rupp, A nonlinear optimal control minimization technique, *Transactions of the American Mathematical Society* 178, 1973, pp. 357–381.

[20] J. Zhang, An augmented Lagrangian approach to simulating yield stress fluid flows around a spherical gas bubble, *International Journal for Numerical Methods in Fluids* 69, 2012, pp. 731-746.