October 2004 **Report LIDS 2632**

Lagrange Multipliers with Optimal Sensitivity Properties in Constrained Optimization[1]

by

Dimitri P. Bertsekas[2]

Abstract

We consider optimization problems with inequality and abstract set constraints, and we derive sensitivity properties of Lagrange multipliers under very weak conditions. In particular, we do not assume uniqueness of a Lagrange multiplier or continuity of the perturbation function. We show that the Lagrange multiplier of minimum norm defines the optimal rate of improvement of the cost per unit constraint violation.

[1] Research supported by NSF Grant ECS-0218328.

[2] Dept. of Electrical Engineering and Computer Science, M.I.T., Cambridge, Mass., 02139.

1. INTRODUCTION

We consider the constrained optimization problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \qquad g_j(x) \le 0, \quad j = 1, \ldots, r, \end{aligned} \tag{P}$$

where X is a nonempty subset of $\Re^n$, and $f : \Re^n \to \Re$ and $g_j : \Re^n \to \Re$ are smooth (continuously differentiable) functions.

In our notation, all vectors are viewed as column vectors, and a prime denotes transposition, so $x'y$ denotes the inner product of the vectors x and y. We will use throughout the standard Euclidean norm $\|x\| = (x'x)^{1/2}$. The gradient vector of a smooth function $h : \Re^n \mapsto \Re$ at a vector x is denoted by $\nabla h(x)$. The positive part of the constraint function $g_j(x)$ is denoted by

$$g_j^+(x) = \max\big\{0, g_j(x)\big\},$$

and we write

$$g(x) = \big(g_1(x), \ldots, g_r(x)\big), \qquad g^+(x) = \big(g_1^+(x), \ldots, g_r^+(x)\big).$$

The *tangent cone* of X at a vector $x \in X$ is denoted by $T_X(x)$. For any cone N, we denote by N^* its polar cone ($N^* = \{z \mid z'y \le 0, \forall\, y \in N\}$). This paper is related to research on Fritz John optimality conditions and associated subjects, described in the papers by Bertsekas and Ozdaglar [BeO02], Bertsekas, Ozdaglar, and Tseng [BOT04], and the book [BNO03]. We generally use the terminology of these works.

A *Lagrange multiplier* associated with a local minimum x^* is a vector $\mu = (\mu_1, \ldots, \mu_r)$ such that

$$\left(\nabla f(x^*) + \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)\right)' d \ge 0, \qquad \forall\, d \in T_X(x^*), \tag{1.1}$$

$$\mu_j \ge 0, \quad \forall\, j = 1, \ldots, r, \qquad \mu_j = 0, \quad \forall\, j \notin A(x^*), \tag{1.2}$$

where $A(x^*) = \big\{j \mid g_j(x^*) = 0\big\}$ is the index set of inequality constraints that are active at x^*. The set of Lagrange multipliers corresponding to x^* is a (possibly empty) closed and convex set.

We will show the following sensitivity result. The proof is given in the next section.

Proposition 1.1: Let x^* be a local minimum of problem (P), assume that the set of Lagrange multipliers is nonempty, and let μ^* be the vector of minimum norm on this set. Then for every sequence $\{x^k\} \subset X$ of infeasible vectors such that $x^k \to x^*$, we have

$$f(x^*) - f(x^k) \leq \|\mu^*\| \|g^+(x^k)\| + o\big(\|x^k - x^*\|\big). \tag{1.3}$$

Furthermore, if $\mu^* \neq 0$ and $T_X(x^*)$ is convex, the preceding inequality is sharp in the sense that there exists a sequence of infeasible vectors $\{x^k\} \subset X$ such that

$$\lim_{k\to\infty} \frac{f(x^*) - f(x^k)}{\|g^+(x^k)\|} = \|\mu^*\|, \tag{1.4}$$

and for this sequence, we have

$$\lim_{k\to\infty} \frac{g_j^+(x^k)}{\|g^+(x^k)\|} = \frac{\mu_j^*}{\|\mu^*\|}, \qquad j = 1, \ldots, r. \tag{1.5}$$

A sensitivity result of this type was first given by Bertsekas, Ozdaglar, and Tseng [BOT04], for the case of a convex, possibly nondifferentiable problem. In that paper, X was assumed convex, and the functions f and g_j were assumed convex over X (rather than smooth). Using the definition of the dual function $[q(\mu) = \inf_{x\in X}\big\{f(x) + \mu' g(x)\big\}]$, it can be seen that

$$q^* - f(x) = q(\mu^*) - f(x) \leq f(x) + \mu^{*\prime} g(x) - f(x) = \mu^{*\prime} g(x) \leq \|\mu^*\|\,\|g^+(x)\|, \qquad \forall\, x \in X,$$

where q^* is the dual optimal value (assumed finite), and μ^* is the dual optimal solution of minimum norm (assuming a dual optimal solution exists). The inequality was shown to be sharp, assuming that $\mu^* \neq 0$, in the sense that there exists a sequence of infeasible vectors $\{x^k\} \subset X$ such that

$$\lim_{k\to\infty} \frac{q^* - f(x^k)}{\|g^+(x^k)\|} = \|\mu^*\|.$$

This result is consistent with Prop. 1.1. However, the line of analysis of the present paper is different, and in fact simpler, because it relies on the machinery of differentiable calculus rather than convex analysis (there is a connection with convex analysis, but it is embodied in Lemma 2.1, given in the next section).

Note that Prop. 1.1 establishes the optimal rate of cost improvement with respect to infeasible constraint perturbations, under much weaker assumptions than earlier results for nonconvex problems. For example, classical sensitivity results, include second order sufficiency assumptions guaranteeing that the Lagrange multiplier is unique and that the *perturbation function*

$$p(u) = \inf_{x\in X,\, g(x)\leq u} f(x)$$

is differentiable (see e.g., [Ber01]). More recent analyses (see, e.g., Bonnans and Shapiro [BoS00], Section 5.2) also require considerably stronger conditions that ours.

Note also that under our weak assumptions, a sensitivity analysis based on the directional derivative of the perturbation function p is not appropriate. The reason is that our assumptions do not preclude the possibility that p has discontinuous directional derivative at $u = 0$, as illustrated by the following example, first discussed in [BOT04].

Example 1.1:

Consider the two-dimensional problem,

$$\begin{aligned} &\text{minimize} && -x_2 \\ &\text{subject to} && x \in X = \{x \mid x_2^2 \le x_1\}, \quad g_1(x) = x_1 \le 0, \quad g_2(x) = x_2 \le 0, \end{aligned}$$

we have

$$p(u) = \begin{cases} -u_2 & \text{if } u_2^2 \le u_1, \\ -\sqrt{u_1} & \text{if } u_1 \le u_2^2,\ u_1 \ge 0,\ u_2 \ge 0, \\ \infty & \text{otherwise.} \end{cases}$$

It can be verified that $x^* = 0$ is the global minimum (in fact the unique feasible solution) and that the set of Lagrange multipliers is

$$\{\mu \ge 0 \mid \mu_2 = 1\}.$$

Consistently with the preceding proposition, for the sequence $x^k = (1/k^2,\, 1/k)$, we have

$$\lim_{k \to \infty} \frac{f(x^*) - f(x^k)}{\|g^+(x^k)\|} = \|\mu^*\| = 1.$$

However, $\mu^* = (0, 1)$, is not a direction of steepest descent, since starting at $u = 0$ and going along the direction $(0, 1)$, $p(u)$ is equal to 0, so

$$p'(0; \mu^*) = 0.$$

In fact p has no direction of steepest descent at $u = 0$, because $p'(0; \cdot)$ is not continuous or even lower semicontinuous. However, one may achieve the optimal improvement rate of $\|\mu^*\|$ by using constraint perturbations that lie on the curved boundary of X.

Finally, let us illustrate with an example how our sensitivity result fails when the convexity assumption on $T_X(x^*)$ is violated. In this connection, it is worth noting that nonconvexity of $T_X(x^*)$ implies that X is not regular at x^* (in the terminology of nonsmooth analysis - see [BNO03] and [RoW78]), and this is a major source of exceptional behavior in relation to Lagrange multipliers (see [BNO03], Chapter 5).

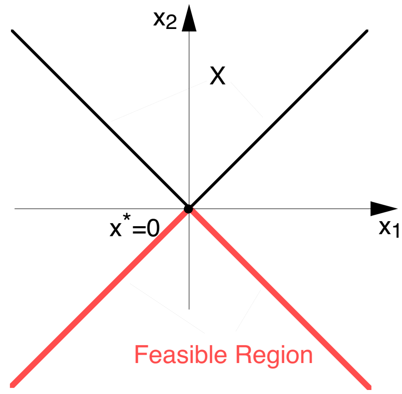

Figure 1.1. Constraints of Example 1.2. We have

$$T_X(x^*) = X = \{x \mid (x_1 + x_2)(x_1 - x_2) = 0\}.$$

The set X consists of the two lines shown, but the feasible region is the lower portion where $x_2 \le 0$.

Example 1.2:

In this 2-dimensional example, there are two linear constraints

$$g_1(x) = x_1 + x_2 \le 0, \qquad g_2(x) = -x_1 + x_2 \le 0,$$

and the set X is the (nonconvex) cone

$$X = \{x \mid (x_1 + x_2)(x_1 - x_2) = 0\}$$

(see Fig. 1.1). Let the cost function be

$$f(x_1, x_2) = x_1^2 + (x_2 - 1)^2.$$

Then the vector $x^* = (0, 0)$ is a local minimum, and we have $T_X(x^*) = X$, so $T_X(x^*)$ is not convex.

A Lagrange multiplier is a nonnegative vector (μ_1^*, μ_2^*) such that

$$\big(\nabla f(x^*) + \mu_1^* \nabla g_1(x^*) + \mu_2^* \nabla g_2(x^*)\big)' d \ge 0, \qquad \forall\ d \in T_X(x^*),$$

from which, since $T_X(x^*)$ contains the vectors $(1, 0)$, $(-1, 0)$, $(0, 1)$, and $(0, -1)$, we obtain

$$\nabla f(x^*) + \mu_1^* \nabla g_1(x^*) + \mu_2^* \nabla g_2(x^*) = 0,$$

or

$$\begin{pmatrix} 0 \\ -2 \end{pmatrix} + \mu_1^* \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mu_2^* \begin{pmatrix} -1 \\ 1 \end{pmatrix} = 0.$$

Thus the unique Lagrange multiplier vector is $\mu^* = (1,1)$. There are two types of sequences $\{x_k\} \subset X$ (and mixtures of these two) that are infeasible and converge to x^*: those that approach x^* along the boundary of the constraint $x_1 + x_2 \le 0$ [these have the form $(-\xi^k, \xi^k)$, where $\xi^k > 0$ and $\xi^k \to 0$], and those that approach x^* along the boundary of the constraint $-x_1 + x_2 \le 0$ [these have the form (ξ^k, ξ^k), where $\xi^k > 0$ and $\xi^k \to 0$]. For any of these sequences, we have $f(x^k) = (\xi^k)^2 + (\xi^k - 1)^2$ and $\|g^+(x^k)\| = 2\xi^k$, so

$$\lim_{k\to\infty} \frac{f(x^*) - f(x^k)}{\|g^+(x^k)\|} = \lim_{k\to\infty} \frac{1 - (\xi^k)^2 - (\xi^k - 1)^2}{2\xi^k} = 1 < \sqrt{2} = \|\mu^*\|.$$

Thus $\|\mu^*\|$ is strictly larger than the optimal rate of cost improvement, and the conclusion of Prop. 1.1 fails.

2. PROOF

Let $\{x^k\} \subset X$ be a sequence of infeasible vectors such that $x^k \to x^*$. We will show the bound (1.3). The sequence $\big\{(x^k - x^*)/\|x^k - x^*\|\big\}$ is bounded and each of its limit points belongs to $T_X(x^*)$. Without loss of generality, we assume that $\big\{(x^k - x^*)/\|x^k - x^*\|\big\}$ converges to a vector $d \in T_X(x^*)$. Then for the minimum norm Lagrange multiplier μ^*, we have

$$\left(\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)' d \ge 0. \tag{2.1}$$

Denote

$$\xi^k = \frac{x^k - x^*}{\|x^k - x^*\|} - d.$$

We have

$$\left(\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)' (x^k - x^*) = \left(\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)' (d + \xi^k)\,\|x^k - x^*\|,$$

so using Eq. (2.1) and the fact $\xi^k \to 0$, we have

$$\left(\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)' (x^k - x^*) \ge o\big(\|x^k - x^*\|\big). \tag{2.2}$$

Using Eq. (2.2), a Taylor expansion, and the fact $\mu^{*\prime} g(x^*) = 0$, we have

$$\begin{aligned} f(x^k) + \mu^{*\prime} g(x^k) &= f(x^*) + \mu^{*\prime} g(x^*) + \left(\nabla f(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)\right)' (x^k - x^*) + o\big(\|x^k - x^*\|\big) \\ &\ge f(x^*) + o\big(\|x^k - x^*\|\big). \end{aligned}$$

We thus obtain, using the fact $\mu \geq 0$,

$$f(x^*) - f(x^k) \leq \mu^{*\prime} g(x^k) + o\big(\|x^k - x^*\|\big) \leq \mu^{*\prime} g^+(x^k) + o\big(\|x^k - x^*\|\big),$$

and using the Cauchy-Schwarz inequality,

$$f(x^*) - f(x^k) \leq \|\mu^*\| \, \|g^+(x^k)\| + o\big(\|x^k - x^*\|\big),$$

which is the desired bound (1.3).

For the proof that the bound is sharp, we will need the following lemma first given in Bertsekas and Ozdaglar [BeO02] (see also [BNO03], Lemma 5.3.1).

Lemma 2.1: Let N be a closed convex cone in $\Re^n$, and let $a_0, \ldots, a_r$ be given vectors in $\Re^n$. Suppose that the set

$$M = \left\{ \mu \geq 0 \;\middle|\; -\left(a_0 + \sum_{j=1}^{r} \mu_j a_j\right) \in N \right\}$$

is nonempty, and let μ^* be the vector of minimum norm in M. Then, there exists a sequence $\{d^k\} \subset N^*$ such that

$$a_0' d^k \to -\|\mu^*\|^2, \qquad (a_j' d^k)^+ \to \mu_j^*, \quad j = 1, \ldots, r. \tag{2.3}$$

For simplicity, we assume that all the constraints are active at x^*. Inactive inequality constraints can be neglected since the subsequent analysis focuses in a small neighborhood of x^*, within which these constraints remain inactive. We will use Lemma 2.1 with the following identifications:

$$N = T_X(x^*)^*, \qquad a_0 = \nabla f(x^*), \qquad a_j = \nabla g_j(x^*), \quad j = 1, \ldots, r,$$

$$M = \text{ set of Lagrange multipliers},$$

$$\mu^* = \text{ Lagrange multiplier of minimum norm}.$$

Since $T_X(x^*)$ is closed and is assumed convex, we have $N^* = T_X(x^*)$, so Lemma 2.1 yields a sequence $\{d^k\} \subset T_X(x^*)$ such that

$$\nabla f(x^*)' d^k \to -\|\mu^*\|^2, \qquad \nabla g_j(x^*)' d^k \to \mu_j^*, \quad j = 1, \ldots, r.$$

Since $d^k \in T_X(x^*)$, for each k we can select a sequence $\{x^{k,t}\} \subset X$ such that $x^{k,t} \neq x^*$ for all t and

$$\lim_{t \to \infty} x^{k,t} = x^*, \qquad \lim_{t \to \infty} \frac{x^{k,t} - x^*}{\|x^{k,t} - x^*\|} = d^k. \tag{2.4}$$

Denote

$$\xi^{k,t} = \frac{x^{k,t} - x^*}{\|x^{k,t} - x^*\|} - d^k.$$

For each k, we select t_k sufficiently large so that

$$\lim_{k\to\infty} \xi^{k,t_k} = 0, \qquad \lim_{k\to\infty} x^{k,t_k} = x^*,$$

and we denote

$$x^k = x^{k,t_k}, \qquad \xi^k = \xi^{k,t_k}.$$

Thus, we have

$$x^k \to x^*, \qquad \xi^k = \frac{x^k - x^*}{\|x^k - x^*\|} - d^k \to 0.$$

Using a first order expansion for the cost function f, we have for each k and t,

$$\begin{aligned} f(x^k) - f(x^*) &= \nabla f(x^*)'(x^k - x^*) + o(\|x^k - x^*\|) \\ &= \nabla f(x^*)' \left(d^k + \xi^k\right) \|x^k - x^*\| + o(\|x^k - x^*\|) \\ &= \|x^k - x^*\| \left(\nabla f(x^*)'d^k + \nabla f(x^*)'\xi^k + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|} \right), \end{aligned}$$

and, since $\xi^k \to 0$ and $\nabla f(x^*)'d^k \to -\|\mu^*\|^2$,

$$f(x^k) - f(x^*) = -\|x^k - x^*\| \cdot \|\mu^*\|^2 + o(\|x^k - x^*\|). \tag{2.5}$$

Similarly, using also the fact $g_j(x^*) = 0$, we have for each k and t,

$$g_j(x^k) = \|x^k - x^*\|\mu_j^* + o(\|x^k - x^*\|), \qquad j = 1, \ldots, r,$$

from which we also have

$$g_j^+(x^k) = \|x^k - x^*\|\mu_j^* + o(\|x^k - x^*\|), \qquad j = 1, \ldots, r. \tag{2.6}$$

We thus obtain

$$\|g^+(x^k)\| = \|x^k - x^*\| \cdot \|\mu^*\| + o(\|x^k - x^*\|), \tag{2.7}$$

which, since $\|\mu^*\| \neq 0$, also shows that $\|g^+(x^k)\| \neq 0$ for all sufficiently large k. Without loss of generality, we assume that

$$\|g^+(x^k)\| \neq 0, \qquad k = 0, 1, \ldots \tag{2.8}$$

By multiplying Eq. (2.7) with $\|\mu^*\|$, we see that

$$\|\mu^*\| \cdot \|g^+(x^k)\| = \|x^k - x^*\|\|\mu^*\|^2 + o(\|x^k - x^*\|). \tag{2.9}$$

Combining Eqs. (2.5) and (2.9), we obtain

$$f(x^*) - f(x^k) = \|\mu^*\| \cdot \|g^+(x^k)\| + o(\|x^k - x^*\|),$$

which together with Eqs. (2.7) and (2.8), shows that

$$\frac{f(x^*) - f(x^k)}{\|g^+(x^k)\|} = \|\mu^*\| + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\| \cdot \|\mu^*\| + o(\|x^k - x^*\|)}.$$

Taking the limit as $k \to \infty$ and using the fact $\|\mu^*\| \neq 0$, we obtain

$$\lim_{k\to\infty} \frac{f(x^*) - f(x^k)}{\|g^+(x^k)\|} = \|\mu^*\|.$$

Finally, from Eqs. (2.6) and (2.7), we see that

$$\frac{g_j^+(x^k)}{\|g^+(x^k)\|} = \frac{\mu_j^*}{\|\mu^*\|} + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|}, \qquad j = 1, \ldots, r,$$

from which Eq. (1.5) follows. **Q.E.D.**

3. REFERENCES

[BNO02] Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., 2003. Convex Analysis and Optimization, Athena Scientific, Belmont, MA.

[BOT04] Bertsekas, D. P., Ozdaglar, A. E., and Tseng, P., 2004. "Enhanced Fritz John Conditions for Convex Programming," Lab. for Information and Decision Systems Report 2631, MIT.

[BeO02] Bertsekas, D. P., and Ozdaglar, A. E., 2002. "Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization," J. Opt. Theory Appl., Vol. 114, 2002, pp. 287-343.

[Ber99] Bertsekas, D. P., 1999. Nonlinear Programming: 2nd Edition, Athena Scientific, Belmont, MA.

[BoS00] Bonnans, J. F., and Shapiro, A., 2000. Perturbation Analysis of Optimization Problems, Springer-Verlag, N. Y.

[Roc70] Rockafellar, R. T., 1970. Convex Analysis, Princeton Univ. Press, Princeton, N. J.

[Roc93] Rockafellar, R. T., 1993. "Lagrange Multipliers and Optimality," SIAM Review, Vol. 35, pp. 183-238.