October 2003

The Relation Between Pseudonormality and Quasiregularity in Constrained Optimization[1]

by

Asuman E. Ozdaglar and Dimitri P. Bertsekas[2]

Abstract

We consider optimization problems with equality, inequality, and abstract set constraints. Our goal is to investigate the connections between various characteristics of the constraint set that imply the existence of Lagrange multipliers. For problems with no abstract set constraint, the classical condition of quasiregularity provides the connecting link between the most common constraint qualifications and existence of Lagrange multipliers. In earlier work, we introduced a new and general condition, pseudonormality, that is central within the theory of constraint qualifications, exact penalty functions, and existence of Lagrange multipliers. In this paper, we explore the relations between pseudonormality, quasiregularity, and existence of Lagrange multipliers in the presence of an abstract set constraint. In particular, under a regularity assumption on the abstract constraint set, we show that pseudonormality implies quasiregularity. However, contrary to pseudonormality, quasiregularity does not imply the existence of Lagrange multipliers, except under additional assumptions.

[1] Research supported by NSF under Grant ACI-9873339.

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

1. INTRODUCTION

In this paper, we consider finite-dimensional optimization problems of the form

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in C, \end{aligned} \tag{1.1}$$

where the constraint set C consists of equality and inequality constraints as well as an additional abstract set constraint $x \in X$:

$$C = X \cap \big\{x \mid h_1(x) = 0, \ldots, h_m(x) = 0\big\} \cap \big\{x \mid g_1(x) \leq 0, \ldots, g_r(x) \leq 0\big\}. \tag{1.2}$$

We assume throughout the paper that f, h_i, g_j are smooth (continuously differentiable) functions from $\Re^n$ to $\Re$, and X is a nonempty closed set.

The analysis of constrained optimization problems is centered around characterizing how the cost function behaves as we move from a local minimum to neighboring feasible points. The relevant variations can be studied in terms of various conical approximations to the constraint set. In this paper, we use two such approximations, namely the *tangent cone* and the *normal cone*, which are particularly useful in characterizing local optimality of feasible solutions of the above problem (see [BNO03]).

In our terminology, a vector y is a *tangent* of a set $S \subset \Re^n$ at a vector $x \in S$ if either $y = 0$ or there exists a sequence $\{x^k\} \subset S$ such that $x^k \neq x$ for all k and

$$x^k \to x, \qquad \frac{x^k - x}{\|x^k - x\|} \to \frac{y}{\|y\|}.$$

The set of all tangent vectors of S at x is denoted by $T_S(x)$ and is also referred to as the *tangent cone* of S at x. The *polar cone* of any cone T is defined by

$$T^* = \big\{z \mid z'y \leq 0,\, y \in T\big\}.$$

The *normal cone* is another conical approximation that is useful in the context of optimality conditions and is of central importance in nonsmooth analysis (see Mordukhovich [Mor76], Aubin and Frankowska [AuF90], Rockafellar and Wets [RoW98], and Borwein and Lewis [BoL00]). For a closed set X and a point $x \in X$, the normal cone of X at x, denoted by $N_X(x)$, is obtained from the polar cone $T_X(x)^*$ by means of a closure operation. In particular, we have $z \in N_X(x)$ if there exist sequences $\{x^k\} \subset X$ and $\{z^k\}$ such that $x^k \to x$, $z^k \to z$, and $z^k \in T_X(x^k)^*$ for all k. From the definition of the normal cone, an important closedness property follows: if $\{x^k\} \subset X$

is a sequence that converges to some $x^* \in X$, and $\{z^k\}$ is a sequence that converges to some z^* with $z^k \in N_X(x^k)$ for all k, then $z^* \in N_X(x^*)$ (see Rockafellar and Wets [RoW98], Proposition 6.6).

It can be seen that, for any $x \in X$, we have $T_X(x)^* \subset N_X(x)$. Note that $N_X(x)$ may not always be equal to $T_X(x)^*$. When $T_X(x)^* = N_X(x)$, we say that X is *regular* at x (see Rockafellar and Wets [RoW98], p. 199). Regularity is an important property, which distinguishes problems that have satisfactory Lagrange multiplier theory from those that do not, as has been been emphasized in earlier work ([Roc93], [BeO00], [BNO03]). In this paper, we further elaborate on the significance of this property of the constraint set in relation to existence of Lagrange multipliers.

A classical necessary condition for a vector $x^* \in C$ to be a local minimum of f over C is

$$\nabla f(x^*)'y \geq 0, \qquad \forall\ y \in T_C(x^*), \tag{1.3}$$

where $T_C(x^*)$ is the tangent cone of C at x^* (see e.g., Bazaraa, Sherali, and Shetty [BSS93], Bertsekas [Ber99], Hestenes [Hes75], Rockafellar [Roc93], Rockafellar and Wets [RoW98]). When the constraint set is specified in terms of equality and inequality constraints, as in problem (1.1)-(1.2), more refined optimality conditions can be obtained. In particular, we say that the constraint set C of Eq. (1.2) *admits Lagrange multipliers* at a point $x^* \in C$ if for every smooth cost function f for which x^* is a local minimum of problem (1.1), there exist vectors $\lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$ and $\mu^* = (\mu_1^*, \ldots, \mu_r^*)$ that satisfy the following conditions:

$$\left(\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right)' y \geq 0, \qquad \forall\ y \in T_X(x^*), \tag{1.4}$$

$$\mu_j^* \geq 0, \qquad \forall\ j = 1, \ldots, r, \tag{1.5}$$

$$\mu_j^* = 0, \qquad \forall\ j \notin A(x^*), \tag{1.6}$$

where $A(x^*) = \big\{j \mid g_j(x^*) = 0\big\}$ is the index set of inequality constraints that are active at x^*. We refer to such a pair (λ^*, μ^*) as a *Lagrange multiplier vector corresponding to f and x^** or simply a *Lagrange multiplier.*

In the case where $X = \Re^n$, a typical approach to asserting the admittance of Lagrange multipliers is to assume structure in the constraint set, which guarantees that the tangent cone $T_C(x^*)$ has the form

$$T_C(x^*) = V(x^*),$$

where $V(x^*)$ is the *cone of first order feasible variations at* x^*, given by

$$V(x^*) = \big\{y \mid \nabla h_i(x^*)'y = 0,\ i = 1, \ldots, m,\ \nabla g_j(x^*)'y \leq 0,\ j \in A(x^*)\big\}. \tag{1.7}$$

In this case we say that x^* is a *quasiregular point* or that *quasiregularity holds at* x^* [other terms used are x^* "satisfies Abadie's constraint qualification" (Abadie [Aba67], Bazaraa, Sherali, and Shetty [BSS93]), or "is a regular point" (Hestenes [Hes75])]. When there is no abstract set constraint, it is well-known (see e.g., Bertsekas [Ber99], p. 332) that for a given smooth f for which x^* is a local minimum, there exist Lagrange multipliers if and only if

$$\nabla f(x^*)'y \geq 0, \qquad \forall\, y \in V(x^*).$$

If x^* is a quasiregular local minimum, it follows from Eq. (1.3) that the above condition holds, so the constraint set admits Lagrange multipliers at x^*. Thus, a common line of analysis when $X = \Re^n$ is to establish various conditions, also known as *constraint qualifications*, which imply quasiregularity, and therefore imply that the constraint set admits Lagrange multipliers (see e.g., Bertsekas [Ber99], or Bazaraa, Sherali, and Shetty [BSS93]). This line of analysis, however, requires fairly complicated proofs to show the relations of constraint qualifications to quasiregularity.

A general constraint qualification, called *quasinormality*, was introduced for the special case where $X = \Re^n$ by Hestenes in [Hes75]. Hestenes also showed that quasinormality implies quasiregularity (see also Bertsekas [Ber99], Proposition 3.3.17). Since it is simple to show that the major classical constraint qualifications imply quasinormality (see e.g. Bertsekas [Ber99]), this provides an alternative line of proof that these constraint qualifications imply quasiregularity. The extension of quasinormality to the case where $X \neq \Re^n$ was investigated in [BeO00]. In particular, we say that a feasible vector x^* of problem (1.1)-(1.2) is *quasinormal* if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$, and no sequence $\{x^k\} \subset X$ such that:

(i) $-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.

(ii) $\mu_j \geq 0$, for all $j = 1, \ldots, r$.

(iii) $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$ are not all equal to 0.

(iv) $\{x^k\}$ converges to x^* and for each k, $\lambda_i h_i(x^k) > 0$ for all i with $\lambda_i \neq 0$ and $\mu_j g_j(x^k) > 0$ for all j with $\mu_j \neq 0$.

A slightly stronger notion, known as *pseudonormality*, was also introduced in [BeO00], and was shown to form the connecting link between major constraint qualifications, admittance of Lagrange multipliers, and exact penalty function theory, for the general case where $X \neq \Re^n$. In particular, we say that a feasible vector x^* of problem (1.1)-(1.2) is *pseudonormal* if there are no scalars $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$, and no sequence $\{x^k\} \subset X$ such that:

(i) $-\left(\sum_{i=1}^m \lambda_i \nabla h_i(x^*) + \sum_{j=1}^r \mu_j \nabla g_j(x^*)\right) \in N_X(x^*)$.

(ii) $\mu_j \geq 0$, for all $j = 1, \ldots, r$, and $\mu_j = 0$ for all $j \notin A(x^*)$.

(iii) $\{x^k\}$ converges to x^* and

$$\sum_{i=1}^{m} \lambda_i h_i(x^k) + \sum_{j=1}^{r} \mu_j g_j(x^k) > 0, \qquad \forall\ k. \tag{1.8}$$

In this paper, we focus on the following extension of the notion of quasiregularity, adopted in several treatments of Lagrange multiplier theory (see e.g. Rockafellar [Roc93]) for the case where X may be a strict subset of $\Re^n$. We say that a feasible vector x^* of problem (1.1)-(1.2) is *quasiregular* if

$$T_C(x^*) = V(x^*) \cap T_X(x^*),$$

where $V(x^*)$ is the cone of first order feasible variations [cf. Eq. (1.7)]. We investigate the connections between pseudonormality, quasiregularity, and admittance of Lagrange multipliers when $X \neq \Re^n$. In Section 2, we show that under a regularity assumption on X, pseudonormality implies quasiregularity. Our line of proof is not only more general than the one of Hestenes (which applies only to the case $X = \Re^n$), but is also considerably simpler. In Section 3, we focus on the relation of quasiregularity and admittance of Lagrange multipliers. We show that contrary to the case where $X = \Re^n$, quasiregularity by itself is not sufficient to guarantee the existence of a Lagrange multiplier. Thus the importance of quasiregularity, which constitutes the classical pathway to Lagrange multipliers when $X = \Re^n$, diminishes when $X \neq \Re^n$. On the other hand, pseudonormality still provides unification of the theory.

Regarding 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 $\|x\| = (x'x)^{1/2}$. We denote the convex hull and the closure of a set C by $\mathrm{conv}(C)$ and $\mathrm{cl}(C)$, respectively.

2. PSEUDONORMALITY, QUASINORMALITY, AND QUASIREGULARITY

In this section, we investigate the connection of pseudonormality and quasiregularity. In particular, we show that under a regularity assumption on X, quasinormality implies quasiregularity even when $X \neq \Re^n$. This shows that any constraint qualification that implies quasinormality also implies quasiregularity. Moreover, since pseudonormality implies quasinormality, it follows that under the given assumption, pseudonormality also implies quasiregularity.

We first give some basic results regarding cones and their polars that will be useful in our analysis. For the proofs, see [Roc70] or [BNO03].

Lemma 1:

(a) Let C be a cone. Then $(C^*)^* = \mathrm{cl}\big(\mathrm{conv}(C)\big)$. In particular, if C is closed and convex, $(C^*)^* = C$.

(b) Let C_1 and C_2 be two cones. If $C_1 \subset C_2$, then $C_2^* \subset C_1^*$.

(c) Let C_1 and C_2 be two cones. Then, $(C_1 + C_2)^* = C_1^* \cap C_2^*$ and $C_1^* + C_2^* \subset (C_1 \cap C_2)^*$. In particular, if C_1 and C_2 are closed and convex, $(C_1 \cap C_2)^* = \mathrm{cl}(C_1^* + C_2^*)$.

Next, we prove the following lemma that relates to the properties of quasinormality. From here on in our analysis, we assume for simplicity that all the constraints of problem (1.1)-(1.2) are inequalities. (Equality constraints can be handled by conversion to inequality constraints.)

Lemma 2: If a vector $x^* \in C$ is quasinormal, then all feasible vectors in a neighborhood of x^* are quasinormal.

Proof: Assume that the claim is not true. Then we can find a sequence $\{x^k\} \subset C$ such that $x^k \neq x^*$ for all k, $x^k \to x^*$ and x^k is not quasinormal for all k. This implies, for each k, the existence of scalars $\xi_1^k, \ldots, \xi_r^k$, and a sequence $\{x_l^k\} \subset X$ such that:

(a)

$$-\left(\sum_{j=1}^{r} \xi_j^k \nabla g_j(x^k)\right) \in N_X(x^k), \tag{2.1}$$

(b) $\xi_j^k \geq 0$, for all $j = 1, \ldots, r$, and $\xi_1^k, \ldots, \xi_r^k$ are not all equal to 0.

(c) $\lim_{l\to\infty} x_l^k = x_k$, and for all l, $\xi_j^k g_j(x_l^k) > 0$ for all j with $\xi_j^k > 0$.

For each k denote

$$\delta^k = \sqrt{\sum_{j=1}^{r} \big(\xi_j^k\big)^2},$$

$$\mu_j^k = \frac{\xi_j^k}{\delta^k}, \qquad j = 1, \ldots, r, \ \forall\ k.$$

Since $\delta^k \neq 0$ and $N_X(x^k)$ is a cone, conditions (a)-(c) for the scalars $\xi_1^k, \ldots, \xi_r^k$ yield the following set of conditions that hold for each k for the scalars $\mu_1^k, \ldots, \mu_r^k$:

(i)

$$-\left(\sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k)\right) \in N_X(x^k), \tag{2.2}$$

(ii) $\mu_j^k \geq 0$, for all $j = 1, \ldots, r$, and $\mu_1^k, \ldots, \mu_r^k$ are not all equal to 0.

(iii) There exists a sequence $\{x_l^k\} \subset X$ such that $\lim_{l\to\infty} x_l^k = x^k$, and for all l, $\mu_j^k g_j(x_l^k) > 0$ for all j with $\mu_j^k > 0$.

Since by construction we have

$$\sum_{j=1}^{r} (\mu_j^k)^2 = 1, \tag{2.3}$$

the sequence $\{\mu_1^k, \ldots, \mu_r^k\}$ is bounded and must contain a subsequence that converges to some nonzero limit $\{\mu_1^*, \ldots, \mu_r^*\}$. Assume without loss of generality that $\{\mu_1^k, \ldots, \mu_r^k\}$ converges to $\{\mu_1^*, \ldots, \mu_r^*\}$. Taking the limit in Eq. (2.2), and using the closedness of the normal cone, we see that this limit must satisfy

$$-\left(\sum_{j=1}^{r} \mu_j^* \nabla g_j(x^*)\right) \in N_X(x^*). \tag{2.4}$$

Moreover, from condition (ii) and Eq. (2.3), it follows that $\mu_j^* \geq 0$, for all $j = 1, \ldots, r$, and $\mu_1^*, \ldots, \mu_r^*$ are not all equal to 0. Finally, let

$$J = \{j \mid \mu_j^* > 0\}.$$

Then, there exists some k_0 such that for all $k \geq k_0$, we must have $\mu_j^k > 0$ for all $j \in J$. From condition (iii), it follows that for each $k \geq k_0$, there exists a sequence $\{x_l^k\} \subset X$ with

$$\lim_{l\to\infty} x_l^k = x^k, \qquad g_j(x_l^k) > 0, \quad \forall\ l,\ \forall\ j \in J.$$

For each $k \geq k_0$, choose an index l_k such that $l_1 < \ldots < l_{k-1} < l_k$ and

$$\lim_{k\to\infty} x_{l_k}^k = x^*.$$

Consider the sequence $\{y^k\}$ defined by

$$y^k = x_{l_{k_0+k-1}}^{k_0+k-1}, \qquad k = 1, 2, \ldots.$$

It follows from the preceding relations that $\{y^k\} \subset X$ and

$$\lim_{k\to\infty} y^k = x^*, \qquad g_j(y^k) > 0, \quad \forall\ k,\ \forall\ j \in J.$$

The existence of scalars $\mu_1^*, \ldots, \mu_r^*$ that satisfy Eq. (2.4) and the sequence $\{y^k\}$ that satisfies the preceding relation violates the quasinormality of x^*, thus completing the proof. **Q.E.D.**

We mentioned that a classical necessary condition for a vector $x^* \in C$ to be a local minimum of the function f over the set C is

$$-\nabla f(x^*) \in T_C(x^*)^*. \tag{2.5}$$

An interesting converse was given by Gould and Tolle [GoT71], namely that every vector in $T_C(x^*)^*$ is equal to the negative of the gradient of some function having x^* as a local minimum over C. Rockafellar and Wets [RoW98, p. 205] showed that this function can be taken to be smooth over $\Re^n$ as in the following lemma.

Lemma 3: Let $\overline{x}$ be a vector in C. For every $y \in T_C(\overline{x})^*$, there is a smooth function F with $-\nabla F(\overline{x}) = y$, which achieves a strict global minimum over C at $\overline{x}$.

We use this result to obtain a specific representation of a vector that belongs to $T_C(\overline{x})^*$ for some $\overline{x} \in C$ under a quasinormality condition, as given in the following proposition. This result will be central in showing the relation of quasinormality to quasiregularity.

Proposition 1: If $\overline{x}$ is a quasinormal vector of C, then any $y \in T_C(\overline{x})^*$ can be represented as

$$y = z + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x}),$$

where $z \in N_X(\overline{x})$, $\overline{\mu}_j \geq 0$, for all $j = 1, \ldots, r$. Furthermore, there exists a sequence $\{x^k\} \subset X$ that converges to $\overline{x}$ and is such that $\overline{\mu}_j g_j(x^k) > 0$ for all k and all j with $\overline{\mu}_j > 0$.

Proof: Let y be a vector that belongs to $T_C(\overline{x})^*$. By Lemma 3, there exists a smooth function F that achieves a strict global minimum over C at $\overline{x}$ with $-\nabla F(\overline{x}) = y$. We use a quadratic penalty function approach. For each $k = 1, 2, \ldots$, choose an $\epsilon > 0$ and consider the "penalized" problem

$$\text{minimize} \;\; F^k(x)$$
$$\text{subject to} \;\; x \in X \cap S,$$

where

$$F^k(x) = F(x) + \frac{k}{2} \sum_{j=1}^{r} \big(g_j^+(x)\big)^2,$$

and $S = \{x \mid ||x - \overline{x}|| \leq \epsilon\}$. Since $X \cap S$ is compact, by Weierstrass' theorem, there exists an optimal solution x^k for the above problem. We have for all k

$$F(x^k) + \frac{k}{2} \sum_{j=1}^{r} \big(g_j^+(x^k)\big)^2 = F^k(x^k) \leq F^k(\overline{x}) = F(\overline{x}) \tag{2.6}$$

and since $F(x^k)$ is bounded over $X \cap S$, we obtain

$$\lim_{k \to \infty} |g_j^+(x^k)| = 0, \quad j = 1, \ldots, r;$$

otherwise the left-hand side of Eq. (2.6) would become unbounded from above as $k \to \infty$. Therefore, every limit point $\tilde{x}$ of $\{x^k\}$ is feasible, i.e., $\tilde{x} \in C$. Furthermore, Eq. (2.6) yields $F(x^k) \le F(\overline{x})$ for all k, so by taking the limit along the relevant subsequence as $k \to \infty$, we obtain

$$F(\tilde{x}) \le F(\overline{x}).$$

Since $\tilde{x}$ is feasible, we have $F(\overline{x}) < F(\tilde{x})$ (since F achieves a strict global minimum over C at $\overline{x}$), unless $\tilde{x} = \overline{x}$, which when combined with the preceding inequality yields $\tilde{x} = \overline{x}$. Thus the sequence $\{x^k\}$ converges to $\overline{x}$, and it follows that x^k is an interior point of the closed sphere S for all k greater than some $\overline{k}$.

For $k \ge \overline{k}$, we have the necessary optimality condition, $\nabla F^k(x^k)'y \ge 0$ for all $y \in T_X(x^k)$, or equivalently $-\nabla F^k(x^k) \in T_X(x^k)^*$, which is written as

$$-\left(\nabla F(x^k) + \sum_{j=1}^{r} \zeta_j^k \nabla g_j(x^k)\right) \in T_X(x^k)^*, \tag{2.7}$$

where

$$\zeta_j^k = k g_j^+(x^k). \tag{2.8}$$

Denote,

$$\delta^k = \sqrt{1 + \sum_{j=1}^{r} (\zeta_j^k)^2}, \tag{2.9}$$

$$\mu_0^k = \frac{1}{\delta^k}, \qquad \mu_j^k = \frac{\zeta_j^k}{\delta^k}, \quad j = 1, \dots, r. \tag{2.10}$$

Then by dividing Eq. (2.7) with δ^k, we get

$$-\left(\mu_0^k \nabla F(x^k) + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k)\right) \in T_X(x^k)^*. \tag{2.11}$$

Since by construction the sequence $\{\mu_0^k, \mu_1^k, \dots, \mu_r^k\}$ is bounded, it must contain a subsequence that converges to some nonzero limit $\{\overline{\mu}_0, \overline{\mu}_1, \dots, \overline{\mu}_r\}$. From Eq. (2.11) and the defining property of the normal cone $N_X(\overline{x})$ [$x^k \to \overline{x},\ z^k \to \overline{z}$, and $z^k \in T_X(x^k)^*$ for all k, imply that $\overline{z} \in N_X(\overline{x})$], we see that $\overline{\mu}_0$ and the $\overline{\mu}_j$ must satisfy

$$-\left(\overline{\mu}_0 \nabla F(\overline{x}) + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x})\right) \in N_X(\overline{x}). \tag{2.12}$$

Furthermore, from Eq. (2.10), we have $g_j(x^k) > 0$ for all j such that $\overline{\mu}_j > 0$ and k sufficiently large. By using the quasinormality of $\overline{x}$, it follows that we cannot have $\overline{\mu}_0 = 0$, and by appropriately normalizing, we can take $\overline{\mu}_0 = 1$ and obtain

$$-\left(\nabla F(\overline{x}) + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x})\right) \in N_X(\overline{x}).$$

Since $-\nabla F(\overline{x}) = y$, we see that

$$y = z + \sum_{j=1}^{r} \overline{\mu}_j \nabla g_j(\overline{x}),$$

where $z \in N_X(\overline{x})$, and the scalars $\overline{\mu}_1, \ldots, \overline{\mu}_r$ and the sequence $\{x^k\}$ satisfy the desired properties, thus completing the proof. **Q.E.D.**

We are now ready to prove the main result of this section.

Proposition 2: If x^* is a quasinormal vector of C and X is regular at x^*, then x^* is quasiregular.

Proof: We must show that $T_C(x^*) = T_X(x^*) \cap V(x^*)$, and to this end, we first show that $T_C(x^*) \subset T_X(x^*) \cap V(x^*)$.

Indeed, since $C \subset X$, using the definition of the tangent cone, we have

$$T_C(x^*) \subset T_X(x^*). \tag{2.13}$$

To show that $T_C(x^*) \subset V(x^*)$, let y be a nonzero tangent of C at x^*. Then there exist sequences $\{\xi^k\}$ and $\{x^k\} \subset C$ such that $x^k \neq x^*$ for all k,

$$\xi^k \to 0, \qquad x^k \to x^*,$$

and

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

By the mean value theorem, we have for all j and k

$$0 \geq g_j(x^k) = g_j(x^*) + \nabla g_j(\tilde{x}^k)'(x^k - x^*) = \nabla g_j(\tilde{x}^k)'(x^k - x^*),$$

where $\tilde{x}^k$ is a vector that lies on the line segment joining x^k and x^* . This relation can be written as

$$\frac{\|x^k - x^*\|}{\|y\|} \nabla g_j(\tilde{x}^k)' y^k \leq 0,$$

where $y^k = y + \xi^k \|y\|$, or equivalently

$$\nabla g_j(\tilde{x}^k)' y^k \leq 0, \qquad y^k = y + \xi^k \|y\|.$$

Taking the limit as $k \to \infty$, we obtain $\nabla g_j(x^*)'y \leq 0$ for all j, thus proving that $y \in V(x^*)$. Hence, $T_C(x^*) \subset V(x^*)$. Together with Eq. (2.13), this shows that

$$T_C(x^*) \subset T_X(x^*) \cap V(x^*). \tag{2.14}$$

To show the reverse inclusion $T_X(x^*) \cap V(x^*) \subset T_C(x^*)$, we first show that

$$N_C(x^*) \subset T_X(x^*)^* + V(x^*)^*.$$

Let y^* be a vector that belongs to $N_C(x^*)$. By the definition of the normal cone, this implies the existence of a sequence $\{x^k\} \subset C$ that converges to x^* and a sequence $\{y^k\}$ that converges to y^*, with $y^k \in T_C(x^k)^*$ for all k. In view of the assumption that x^* is quasinormal, it follows from Lemma 2 that for all sufficiently large k, x^k is quasinormal. Then, by Proposition 1, for each sufficiently large k, there exists a vector $z^k \in N_X(x^k)$ and nonnegative scalars $\mu_1^k, \ldots, \mu_r^k$ such that

$$y^k = z^k + \sum_{j=1}^{r} \mu_j^k \nabla g_j(x^k). \tag{2.15}$$

Furthermore, there exists a sequence $\{x_l^k\} \subset X$ such that

$$\lim_{l \to \infty} x_l^k = x^k,$$

and for all l, $\mu_j^k g_j(x_l^k) > 0$ for all j with $\mu_j^k > 0$.

We will show that the sequence $\{\mu_1^k, \ldots, \mu_r^k\}$ is bounded. Suppose, to arrive at a contradiction, that this sequence is unbounded, and assume without loss of generality, that for each k, at least one of the μ_j^k is nonzero. For each k, denote

$$\delta^k = \frac{1}{\sum_{j=1}^{r} (\mu_j^k)^2},$$

and

$$\xi_j^k = \delta^k \mu_j^k, \qquad \forall\ j = 1, \ldots, r.$$

It follows that $\delta^k > 0$ for all k and $\delta^k \to 0$ as $k \to \infty$. Then, by multiplying Eq. (2.15) by δ^k, we obtain

$$\delta^k y^k = \delta^k z^k + \sum_{j=1}^{r} \xi_j^k \nabla g_j(x^k),$$

or equivalently, since $z^k \in N_X(x^k)$ and $\delta^k > 0$, we have

$$\delta^k z^k = \left(\delta^k y^k - \sum_{j=1}^{r} \xi_j^k \nabla g_j(x^k) \right) \in N_X(x^k).$$

Note that by construction, the sequence $\{\xi_1^k, \ldots, \xi_r^k\}$ is bounded, and therefore has a nonzero limit point $\{\xi_1^*, \ldots, \xi_r^*\}$. Taking the limit in the preceding relation along the relevant subsequence and using the facts $\delta^k \to 0$, $y^k \to y^*$, and $x^k \to x^*$ together with the closedness of the normal cone $N_X(x^*)$, we see that $\delta^k z^k$ converges to some vector z^* in $N_X(x^*)$, where

$$z^* = -\left(\sum_{j=1}^{r} \xi_j^* \nabla g_j(x^*) \right).$$

Furthermore, by defining an index l_k for each k such that $l_1 < \cdots < l_{k-1} < l_k$ and

$$\lim_{k\to\infty} x^k_{l_k} = x^*,$$

we see that for all j with $\xi^*_j > 0$, we have $g_j(x^k_{l_k}) > 0$ for all sufficiently large k. The existence of such scalars $\xi^*_1, \ldots, \xi^*_r$ violates the quasinormality of the vector x^*, thus showing that the sequence $\{\mu^k_1, \ldots, \mu^k_r\}$ is bounded.

Let $\{\mu^*_1, \ldots, \mu^*_r\}$ be a limit point of the sequence $\{\mu^k_1, \ldots, \mu^k_r\}$, and assume without loss of generality that $\{\mu^k_1, \ldots, \mu^k_r\}$ converges to $\{\mu^*_1, \ldots, \mu^*_r\}$. Taking the limit as $k \to \infty$ in Eq. (2.15), we see that z^k converges to some z^*, where

$$z^* = y^* - \left(\sum_{j=1}^r \mu^*_j \nabla g_j(x^*)\right). \tag{2.16}$$

By closedness of the normal cone $N_X(x^*)$ and in view of the assumption that X is regular at x^*, so that $N_X(x^*) = T_X(x^*)^*$, we have that $z^* \in T_X(x^*)^*$. Furthermore, by defining an index l_k for each k such that $l_1 < \cdots < l_{k-1} < l_k$ and

$$\lim_{k\to\infty} x^k_{l_k} = x^*,$$

we see that for all j with $\mu^*_j > 0$, we have $g_j(x^k_{i_k}) > 0$ for all sufficiently large k, showing that $g_j(x^*) = 0$. Hence, it follows that $\mu^*_j = 0$ for all $j \notin A(x^*)$, and using Eq. (2.16), we can write y^* as

$$y^* = z^* + \left(\sum_{j\in A(x^*)} \mu^*_j \nabla g_j(x^*)\right).$$

By Farkas' Lemma, $V(x^*)^*$ is the cone generated by $\nabla g_j(x^*),\ j \in A(x^*)$. Hence, it follows that $y^* \in T_X(x^*)^* + V(x^*)^*$, and we conclude that

$$N_C(x^*) \subset T_X(x^*)^* + V(x^*)^*. \tag{2.17}$$

Finally, using the properties relating to cones and their polars given in Lemma 1 and the fact that $T_X(x^*)$ is convex (which follows by the regularity of X at x^*, see [Row98, p. 221]), we obtain

$$\left(T_X(x^*)^* + V(x^*)^*\right)^* = T_X(x^*) \cap V(x^*) \ \subset\ N_C(x^*)^*. \tag{2.18}$$

Using the relation $N_C(x^*)^* \subset T_C(x^*)$ (see [Row98], Propositions 6.26 and 6.28), this shows that $T_X(x^*) \cap V(x^*) \subset T_C(x^*)$, which together with Eq. (2.14) concludes the proof. **Q.E.D.**

Note that in the preceding proof, we showed that

$$T_C(x^*) \subset T_X(x^*) \cap V(x^*),$$

which, using the relations given in Lemma 1, implies that

$$T_X(x^*)^* + V(x^*)^* \subset \big(T_X(x^*) \cap V(x^*)\big)^* \subset T_C(x^*)^*.$$

We also proved that if X is regular at x^* and x^* is quasinormal, we have

$$N_C(x^*) \subset T_X(x^*)^* + V(x^*)^*,$$

[cf. Eq. (2.17)]. Combining the preceding two relations with the relation $T_C(x^*)^* \subset N_C(x^*)$, we obtain

$$T_C(x^*)^* = N_C(x^*),$$

thus showing that quasinormality of x^* together with regularity of X at x^* implies that C is regular at x^*.

3. QUASIREGULARITY AND ADMITTANCE OF LAGRANGE MULTIPLIERS

In this section, we explore the relation of quasiregularity with the admittance of Lagrange multipliers. First, we provide a necessary and sufficient condition for the constraint set C of Eq. (1.2) to admit Lagrange multipliers. This condition was given in various forms by Gould and Tolle [GoT72], Guignard [Gui69], and Rockafellar [Roc93]. For completeness, we provide the proof of this result using the gradient characterization of vectors in $T_C(x^*)^*$, given in Lemma 3.

Proposition 3: Let x^* be a feasible vector of problem (1.1)-(1.2). The constraint set C of problem (1.2) admits Lagrange multipliers at x^* if and only if

$$T_C(x^*)^* = T_X(x^*)^* + V(x^*)^*. \tag{3.1}$$

Proof: Denote by $D(x^*)$ the set of gradients of all smooth cost functions for which x^* is a local minimum of problem (1.1)-(1.2). We claim that $D(x^*) = T_C(x^*)$. Indeed by the necessary condition for optimality [cf. Eq. (1.3)], we have

$$-D(x^*) \subset T_C(x^*)^*.$$

To show the reverse inclusion, let $y \in T_C(x^*)^*$. By Lemma 3, there exists a smooth function F with $-\nabla F(x^*) = y$, which achieves a strict global minimum over C at $\overline{x}$. Thus, $y \in -D(x^*)$, showing that

$$-D(x^*) = T_C(x^*)^*. \tag{3.2}$$

We now note that by definition, the constraint set C admits Lagrange multipliers at x^* if and only if

$$-D(x^*) \subset T_X(x^*)^* + V(x^*)^*.$$

In view of Eq. (3.2), this implies that the constraint set C admits Lagrange multipliers at x^* if and only if

$$T_C(x^*)^* \subset T_X(x^*)^* + V(x^*)^*. \tag{3.3}$$

On the other hand, we have shown in the proof of Proposition 2 that we have

$$T_C(x^*) \subset T_X(x^*) \cap V(x^*),$$

[cf. Eq. (2.14)]. Using the properties of polar cones given in Lemma 1, this implies

$$T_X(x^*)^* + V(x^*)^* \subset \big(T_X(x^*) \cap V(x^*)\big)^* \subset T_C(x^*)^*,$$

which combined with Eq. (3.3), yields the desired relation, and concludes the proof. **Q.E.D.**

Note that the condition given in Eq. (3.1) is equivalent to the following two conditions:

(a) $V(x^*) \cap \mathrm{cl}\Big(\mathrm{conv}\big(T_X(x^*)\big)\Big) = \mathrm{cl}\Big(\mathrm{conv}\big(T_C(x^*)\big)\Big)$,

(b) $V(x^*)^* + T_X(x^*)^*$ is a closed set.

Quasiregularity is a weaker condition, even under the assumption that X is regular, since the vector sum $V(x^*)^* + T_X(x^*)^*$ need not be closed even if both of these cones themselves are closed, as shown in the following example.

Example 1:

Consider the constraint set $C \subset \Re^3$ specified by

$$C = \big\{x \in X \mid h(x) = 0\big\},$$

where

$$X = \big\{(x_1, x_2, x_3) \mid x_1^2 + x_2^2 \le x_3^2,\ x_3 \le 0\big\},$$

and

$$h(x) = x_2 + x_3.$$

Let x^* denote the origin. In view of the convexity of X, we have that X is regular at x^* and that $T_X(x^*)$ is given by the closure of the set of feasible directions at x^*. Since x^* is the origin and X is a closed cone, it follows that

$$T_X(x^*) = X.$$

The cone of first order feasible variations, $V(x^*)$, is given by

$$V(x^*) = \{(x_1, x_2, x_3) \mid x_2 + x_3 = 0\}.$$

It can be seen that the set $V(x^*)^* + T_X(x^*)^*$ is not closed (cf. [BNO03], Exercise 3.5), implying that C does not admit Lagrange multipliers. On the other hand, we have

$$T_C(x^*) = T_X(x^*) \cap V(x^*),$$

i.e., x^* is quasiregular.

The preceding example shows that quasiregularity is not powerful enough to assert the existence of Lagrange multipliers for the general case $X \neq \Re^n$, unless additional assumptions are imposed. It is effective only for special cases, for instance, when $T_X(x^*)$ is a polyhedral cone, in which case $V(x^*)^* + T_X(x^*)^*$ is closed, since it is the vector sum of two polyhedral sets, and quasiregularity implies the admittance of Lagrange multipliers. Thus the importance of quasiregularity, the classical pathway to Lagrange multipliers when $X = \Re^n$, diminishes when $X \neq \Re^n$. By contrast, as shown in [BeO00], pseudonormality still provides unification of the theory of constraint qualifications and existence of Lagrange multipliers.

REFERENCES

[Aba67] Abadie, J., 1967. "On the Kuhn-Tucker Theorem," in Nonlinear Programming, Abadie, J., (Ed.), North Holland, Amsterdam.

[AuF90] Aubin, J.-P., and Frankowska, H., Set-Valued Analysis, Birkhauser, Boston.

[BSS93] Bazaraa, M. S., Sherali, H. D., and Shetty, C. M., 1993. Nonlinear Programming Theory and Algorithms, (2nd Ed.), Wiley, N. Y.

[BaG82] Bazaraa, M. S., and Goode, J. J., 1982. "Sufficient Conditions for a Globally Exact Penalty Function without Convexity," Math. Programming Stud., Vol. 19, pp. 1-15.

[BNO03] Bertsekas, D. P., Nedic, A., and Ozdaglar, A. E., Convex Analysis and Optimization, Athena Scientific, Belmont, MA, 2003.

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

[BeO00] Bertsekas, D. P., and Ozdaglar, A. E., "Pseudonormality and Lagrange Multiplier Theory for Constrained Optimization," Report LIDS-P-2489, Lab. for Information and Decision Systems, MIT, Nov. 2000, JOTA, to appear.

[BoL00] Borwein, J. M., and Lewis, A. S., Convex Analysis and Nonlinear Optimization, Springer-Verlag, N. Y.

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

[GoT71] Gould, F. J., and Tolle, J., 1971. "A Necessary and Sufficient Condition for Constrained Optimization," SIAM J. Applied Math., Vol. 20, pp. 164-172.

[GoT72] Gould, F. J., and Tolle, J., 1972. "Geometry of Optimality Conditions and Constraint Qualifications," Math. Programming, Vol. 2, pp. 1-18.

[Gui69] Guignard, M., 1969. "Generalized Kuhn-Tucker Conditions for Mathematical Programming Problems in a Banach Space," SIAM J. on Control, Vol. 7, pp. 232-241.

[Hes75] Hestenes, M. R., 1975. Optimization Theory: The Finite Dimensional Case, Wiley, N. Y.

[Mor76] Mordukhovich, B. S., 1976. "Maximum Principle in the Problem of Time Optimal Response with Nonsmooth Constraints," J. of Applied Mathematics and Mechanics, Vol. 40, pp. 960-969.

[RoW98] Rockafellar, R. T., and Wets, R. J.-B., 1998. Variational Analysis, Springer-Verlag, Berlin.

[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.