SIAM J. NUMER. ANAL.
Vol. 25, No. 5, October 1988

014

ON THE IDENTIFICATION OF ACTIVE CONSTRAINTS*

JAMES V. BURKE† AND JORGE J. MORÉ‡

Abstract. Nondegeneracy conditions that guarantee that the optimal active constraints are identified in a finite number of iterations are studied. Results of this type have only been established for a few algorithms, and then under restrictive hypothesis. The main result is a characterization of those algorithms that identify the optimal constraints in a finite number of iterations. This result is obtained with a nondegeneracy assumption which is equivalent, in the standard nonlinear programming problem, to the assumption that there is a set of strictly complementary Lagrange multipliers. As an important consequence of the authors' results the way that this characterization applies to gradient projection and sequential quadratic programming algorithms is shown.

Key words. projected gradient, nondegeneracy, strict complementarity, active constraints, constrained optimization

AMS(MOS) subject classifications. 65K05, 90C20, 90C25, 90C30

1. Introduction. The problem of minimizing a continuously differentiable mapping $f: R^n \to R$ over a nonempty closed convex set $\Omega \subset R^n$,

$$\min \{f(x): x \in \Omega\}, \tag{1.1}$$

has received considerable attention. The linearly constrained case where Ω is a polyhedron is of special interest. Our discussion centers on certain geometrical aspects that arise in the analysis of algorithms for the numerical solution of problem (1.1). We are interested in nondegeneracy conditions that guarantee that the optimal active constraints are identified in a finite number of iterations. Results of this type have only been established for a few algorithms, and then under restrictive hypothesis. Our main result is a characterization of those algorithms that identify the optimal constraints in a finite number of iterations. This result is obtained with a nondegeneracy assumption which is equivalent, in the standard nonlinear programming problem, to the assumption that there is a set of strictly complementary Lagrange multipliers. As an important consequence of our results, we show that this characterization applies to gradient projection algorithms and to sequential quadratic programming algorithms.

Motivation for this work came from recent convergence results on the gradient projection algorithm. Given an inner product norm $\|\cdot\|$, the *projection* into a nonempty closed convex set Ω is the mapping $P: R^n \to \Omega$ defined by

$$P(x) = \operatorname{argmin} \{\|z - x\|: z \in \Omega\}. \tag{1.2}$$

The *gradient projection* algorithm is defined by

$$x_{k+1} = P(x_k - \alpha_k \nabla f(x_k)), \tag{1.3}$$

where $\alpha_k > 0$ is the step and ∇f is the gradient of f with respect to the inner product

* Received by the editors November 6, 1986; accepted for publication (in revised form) October 9, 1987.

† Department of Mathematics, GN-50, University of Washington, Seattle, Washington 98195. The work of this author was supported in part by the National Science Foundation under grant DMS-8602399 and by the Air Force Office of Scientific Research under grant ISSA-860080, and by the Applied Mathematical Sciences subprogram of the Office of Energy Research of the U.S. Department of Energy under contract W-31-109-Eng-38.

‡ Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois 60439. The work of this author was supported in part by the Applied Mathematical Sciences subprogram of the Office of Energy Research of the U.S. Department of Energy under contract W-31-109-Eng-38.

associated with the norm $\|\cdot\|$. The dependence of P on Ω is usually clear from the context, but if there is a possibility of confusion P_Ω will be used to denote the projection into Ω.

An interesting aspect of Dunn's [1987] work on the gradient projection algorithm is that his definition of nondegeneracy is geometric. In particular, the definition of nondegeneracy is independent of the representation of Ω by constraints, and is valid for any convex set Ω. Moreover, this definition of nondegeneracy is weaker than the standard definition which requires independence of the active constraint normals and strictly complementary Lagrange multipliers. Another interesting aspect of Dunn's work is that it is not necessary to assume that Ω is a polyhedron. This is done by replacing the notion of active constraints by the notion of an open facet of Ω, and showing that the gradient projection algorithm identifies the optimal facet in a finite number of iterations.

Calamai and Moré [1987] introduced the notion of the projected gradient $\nabla_\Omega f$, and showed that if $\{x_k\}$ is the sequence generated by the gradient projection algorithm then $\{\nabla_\Omega f(x_k)\}$ converges to zero. This result strengthened previous convergence results (see, for example, Bertsekas [1976] and Dunn [1981]) because it implies that any limit point of $\{x_k\}$ is a stationary point for problem (1.1). In the linearly constrained case, Calamai and Moré also showed that if $\{x_k\}$ is any sequence which converges to a stationary point x^* that is nondegenerate in the classical sense, and if $\{\nabla_\Omega f(x_k)\}$ converges to zero, then the active constraints at x^* are identified in a finite number of iterations. This result is of interest because it applies to any algorithm that generates a sequence $\{x_k\}$ such that $\{\nabla_\Omega f(x_k)\}$ converges to zero.

The concept of the face of a convex set is important to our development because the face of a polyhedron is the geometric analogue of the set of active constraints. We thus begin our development by discussing the geometry of faces. For general convex sets, our results hold if we consider the class of quasi-polyhedral faces defined in § 2. We provide a topological characterization of quasi-polyhedral faces, and in particular, we show that every face of a polyhedron is quasi-polyhedral. It is also shown that quasi-polyhedral faces are essentially the open facets of Dunn [1987]. Quasi-polyhedral faces, however, have a more transparent definition.

We consider the nondegeneracy condition of Dunn [1987] in § 3, and show that this condition is a geometric generalization of the standard strict complementarity condition. For linearly constrained problems and for convex sets defined by constraint functions that satisfy a constraint qualification, this nondegeneracy condition is shown to be equivalent to the assumption that there is a set of strictly complementary Lagrange multipliers; linear independence of the active constraint normals is not needed. We also characterize those algorithms that achieve the optimal face in a finite number of iterations. We prove that if $\{x_k\}$ is any sequence that converges to a nondegenerate stationary point x^* in the relative interior of a quasi-polyhedral face, then the optimal face is identified in a finite number of iterations if and only if $\{\nabla_\Omega f(x_k)\}$ converges to zero. This characterization generalizes and extends the results of Calamai and Moré [1987], and Dunn [1987].

We consider sequential quadratic programming algorithms and gradient projection methods in § 4, and show how the characterization result applies to these algorithms. In particular, we show that the sequential quadratic programming algorithm for linearly constrained problems identifies the optimal active constraints if the limit point of the iterates is nondegenerate in the sense of § 3.

We finish the paper by establishing a connection between the projected gradient of f and the Clarke subdifferential of the composite mapping $f \circ P$. This relationship is of interest because it relates our results to the convergence results for nondifferentiable optimization.

2. Face geometry. We assume a basic background in convex analysis. In particular, recall that for a set $S \subset R^n$ the *affine hull* aff (S) is the smallest affine set which contains S, and the *relative interior* ri (S) is the interior of S relative to aff (S).

Let Ω be a convex set in R^n. A convex set $\Omega_F \subset \Omega$ is a *face* of Ω if the endpoints of any closed line segment in Ω whose relative interior intersects Ω_F are contained in Ω_F. Thus, if x and y are in Ω and $\lambda x + (1-\lambda)y$ lies in Ω_F for some $0 < \lambda < 1$, then x and y must also belong to Ω_F. This terminology is fairly standard (see, for example, Rockafellar [1970]), although other authors use the term *extreme subset* and reserve the term face for the extreme subsets of a polyhedron. A set $\Omega_F \subset \Omega$ is *exposed* if

$$\Omega_F = \operatorname{argmax}\{\phi(x) \colon x \in \Omega\}, \tag{2.1}$$

for some linear functional $\phi : R^n \to R$. It is not difficult to show that (2.1) implies that Ω_F is a face, so the term *exposed face* is justified. Also note that not all faces are exposed. For example, if

$$\Omega = \{(\xi_1, \xi_2) \colon \xi_2 \leqq (1-\xi_1^2)^{1/2}, 0 \leqq \xi_1 \leqq 1\}, \tag{2.2}$$

then $(1, 0)$ is a face of Ω, but it is not an exposed face.

The key to understanding the geometry of the faces of a convex set lies in the following two well-known results (see, for example, Theorems 18.1 and 18.2 of Rockafellar [1970]).

THEOREM 2.1. *Let Ω_F be a face of the convex set Ω. If Γ is a convex subset of Ω such that* ri (Γ) *meets Ω_F, then $\Gamma \subset \Omega_F$.*

This result shows that it is possible to replace line segments by convex subsets in the definition of face. Also note that it implies that the relative interior of distinct faces do not intersect. The next result shows, in particular, that each $x \in \Omega$ can be associated with a unique face of Ω.

THEOREM 2.2. *The collection of all relative interiors of faces is a partition of a convex set Ω.*

We will use the notation $\Omega(x)$ to denote the unique face of Ω such that $x \in$ ri $(\Omega(x))$. For example, if Ω is defined by (2.2) then $\Omega(1, \xi_2)$ is independent of ξ_2 for $\xi_2 < 0$, but a new face is obtained for $\xi_2 = 0$. Note that for this Ω there are an infinite number of faces of the form $\Omega(x)$.

We now show that the tangent and normal cones at points x in the relative interior of a face are independent of x. Recall that for a convex set Ω and a point x in Ω, the *normal cone* at x is defined by

$$N(x) = \{u \in R^n \colon (u, y - x) \leqq 0, y \in \Omega\}.$$

The tangent cone can be defined by polarity. The *polar* K° of a cone K in R^n is the set of all $u \in R^n$ such that $(u, v) \leqq 0$ for all $v \in K$. The *tangent cone* $T(x)$ at x is then $N(x)^\circ$.

THEOREM 2.3. *If Ω_F is a face of the convex set Ω, then $N(x)$ is independent of x for $x \in$* ri (Ω_F).

Proof. Let x_1 and x_2 belong to ri (Ω_F). Choose $u \in N(x_1)$ and consider the exposed face

$$\Omega_E \equiv \operatorname{argmax}\{(u, x) \colon x \in \Omega\}.$$

Clearly, $x_1 \in \Omega_E$ since $u \in N(x_1)$, and thus ri (Ω_F) meets Ω_E. Theorem 2.1 thus shows that $\Omega_F \subset \Omega_E$. In particular, $x_2 \in \Omega_E$, and thus $u \in N(x_2)$. This argument shows that

$N(x_1) \subset N(x_2)$. Since the argument is symmetric in x_1 and x_2, this establishes the result. □

From Theorem 2.3 we see that it makes sense to speak of the normal cone and the tangent cone of a face of Ω. Consequently, we make the following definition.

DEFINITION 2.4. If Ω_F is a face of the convex set Ω, then the normal cone $N(\Omega_F)$ and the tangent cone $T(\Omega_F)$ are, respectively, the normal cone and the tangent cone at any $x \in \text{ri}(\Omega_F)$.

As opposed to general convex sets, the face structure of a polyhedron is more tractable. For example, every face of a polyhedron is exposed, while this is not the case for general convex sets. Faces of polyhedrons have several other properties that are not shared by faces in general. The property needed for our purposes can be expressed in terms of the linearity of the tangent cone of the face; for a cone $K \subset R^n$, the *linearity* $\text{lin}\{K\}$ of the cone is the largest subspace contained in K.

DEFINITION 2.5. A face Ω_F of a convex set Ω is quasi-polyhedral if

$$\text{aff}(\Omega_F) = x + \text{lin}\{T(x)\} \tag{2.3}$$

for any $x \in \text{ri}(\Omega_F)$.

The relative interior of a convex set is a quasi-polyhedral face. As another example, note that if Ω is the convex set defined by (2.2) then there are four quasi-polyhedral faces of the form $\Omega(x)$: the relative interior of Ω, the line segments $\Omega(1, -1)$ and $\Omega(0, 0)$, and the point $\Omega(0, 1)$. Note that the face $\Omega(1, 0)$ is not quasi-polyhedral.

We only require that (2.3) hold for $x \in \text{ri}(\Omega_F)$. This is sufficient for our purposes because Theorem 2.2 guarantees that any $x \in \Omega$ lies in the relative interior of a unique face Ω_F. Also note that if $x \in \text{ri}(\Omega_F)$ then $\text{aff}(\Omega_F) - x$ is a subspace in $T(x)$, and thus

$$\text{aff}(\Omega_F) \subset x + \text{lin}\{T(x)\}. \tag{2.4}$$

This is not difficult to verify because if $z \in \text{aff}(\Omega_F)$ then $x + \mu(z - x)$ is in $\text{aff}(\Omega_F)$ for all scalars μ, and since $x \in \text{ri}(\Omega_F)$, we must have $x + \mu(z - x)$ in Ω_F if $|\mu|$ is sufficiently small. In particular, z belongs to $x + T(x)$ as desired. We now show that the reverse inclusion holds if Ω is the polyhedron defined by

$$\Omega = \{x \in R^n : (c_j, x) \geqq \delta_j, j = 1, \cdots, m\} \tag{2.5}$$

for some vectors $c_j \in R^n$ and scalars δ_j. For future reference, we define

$$A(x) \equiv \{j : (c_j, x) = \delta_j\}$$

as the set of *active constraints at x*.

THEOREM 2.6. *If Ω_F is a face of the polyhedron Ω then Ω_F is a quasi-polyhedral face.*

Proof. As noted above, we only need to show that if $x \in \text{ri}(\Omega_F)$ then

$$x + \text{lin}\{T(x)\} \subset \text{aff}(\Omega_F).$$

If Ω is the polyhedron (2.5), then a short computation verifies that

$$\text{lin}\{T(x)\} = \{\nu \in R^n : (c_j, \nu) = 0, j \in A(x)\}. \tag{2.6}$$

Now choose y in $x + \text{lin}\{T(x)\}$. Then (2.6) shows that if $|\mu|$ is sufficiently small, then

$$x_\mu \equiv x + \mu(y - x) \in \Omega.$$

Since $x \in \Omega_F$, the definition of a face implies that $x_\mu \in \Omega_F$. Since $y \in \text{aff}\{x, x_\mu\}$, we obtain that $y \in \text{aff}(\Omega_F)$ as desired. □

Theorems 2.2 and 2.6 show that any x in a polyhedron Ω lies in the relative interior of a quasi-polyhedral face of Ω. Given a set of constraints A of the polyhedron (2.5), the relative interior of a quasi-polyhedral face is defined by

$$\text{ri}(\Omega_F)=\{x\in R^n: (c_j,x)=\delta_j, j\in A, (c_j,x)>\delta_j, j\notin A\}. \tag{2.7}$$

Conversely, given a quasi-polyhedral face Ω_F of the polyhedron (2.5) the set of active constraints $A(x)$ is independent of $x\in\text{ri}(\Omega_F)$ and satisfies (2.7) with $A=A(x)$.

A quasi-polyhedral face need not be polyhedral, and conversely, a polyhedral face need not be quasi-polyhedral. For example, the bases of a right circular cylinder are quasi-polyhedral faces but are not polyhedral. Also, all boundary line segments that connect the bases in a perpendicular fashion are polyhedral faces, but are not quasi-polyhedral.

Quasi-polyhedral faces are closely related to the concept of an open facet as defined by Dunn [1987]. Open facets are defined in terms of the orthogonal complement of the normal cone: for a set $S\subset R^n$ the *orthogonal complement* of S is the subspace $S^\perp$ of all vectors orthogonal to S. A nonempty subset Ω_F of Ω is an *open facet* of Ω if there is an affine subspace V such that

$$V=x+N(x)^\perp \tag{2.8}$$

for all $x\in\Omega_F$, and

$$\Omega_F=\text{int}_V(\Omega\cap V),$$

where $\text{int}_V(\cdot)$ is the interior with respect to V. We now show that any open facet is the relative interior of a quasi-polyhedral face, and that the relative interior of a quasi-polyhedral face is an open facet. The following result is the first step in establishing this relationship.

LEMMA 2.7. *Let K be a closed convex cone. Then*

$$\text{lin}\{K\}=\{K^\circ\}^\perp.$$

Proof. Since K is a closed convex cone, $K^{\circ\circ}=K$. Hence,

$$[K^\circ]^\perp\subset K^{\circ\circ}=K,$$

and since $[K^\circ]^\perp$ is a subspace,

$$[K^\circ]^\perp\subset\text{lin}\{K\}.$$

Conversely, let $x\in\text{lin}\{K\}$ and choose any $\nu\in K^\circ$. Then $(x,\nu)\leqq 0$, and since $-x\in\text{lin}\{K\}$ also, $(x,\nu)\geqq 0$. Hence, $(x,\nu)=0$, and this implies that $x\in[K^\circ]^\perp$. Thus,

$$\text{lin}\{K\}\subset[K^\circ]^\perp$$

as desired. □

Lemma 2.7 shows that if Ω_F is a quasi-polyhedral face then $\text{aff}(\Omega_F)=V$ where V is defined by (2.8) for any $x\in\text{ri}(\Omega_F)$. Since Ω_F is a face, a computation based on Theorem 2.1 shows that

$$\Omega_F=\Omega\cap\text{aff}(\Omega_F).$$

Thus $\Omega_F=\Omega\cap V$, and this implies that the relative interior of Ω_F is an open facet.

We now show that an open facet Ω_F is the relative interior of a quasi-polyhedral face. We begin by proving that $\text{aff}(\Omega_F)=V$. It is clear that $\text{aff}(\Omega_F)$ is in V. Now let $x\in\Omega_F$ and choose any $y\in V$. Then $x_\mu\equiv x+\mu(y-x)$ belongs to V, and since $\Omega_F=\text{int}_V(\Omega\cap V)$, we must have $x_\mu\in\Omega_F$ for all μ sufficiently small. Since $y\in\text{aff}\{x,x_\mu\}$, we obtain that $y\in\text{aff}(\Omega_F)$ as desired. This proves, in particular, that

$$\Omega_F=\text{ri}(\Omega_F).$$

We complete the proof by showing that if $x \in \Omega_F$ and $\Omega(x)$ is the face of Ω with $x \in \text{ri}(\Omega(x))$, then $\Omega_F \in \text{ri}(\Omega(x))$ and $\Omega(x)$ is quasi-polyhedral. Theorem 2.2 guarantees the existence of $\Omega(x)$. Also note that since Ω_F is convex and $x \in \Omega_F = \text{ri}(\Omega_F)$, Theorem 2.1 shows that $\Omega_F \subset \Omega(x)$. Thus $\text{aff}(\Omega_F) \subset \text{aff}(\Omega(x))$, and since (2.4) holds for any $x \in \text{ri}(\Omega_F)$,

$$\text{aff}(\Omega(x)) \subset x + \text{lin}\{T(x)\} = \text{aff}(\Omega_F).$$

This proves that $\Omega(x)$ is a quasi-polyhedral face and that $\text{aff}(\Omega(x)) = \text{aff}(\Omega_F) = V$. Hence

$$\text{ri}(\Omega(x)) \subset \text{int}_V(\Omega \cap V) = \Omega_F,$$

and since $\Omega_F \subset \Omega(x)$,

$$\Omega_F = \text{ri}(\Omega_F) \subset \text{ri}(\Omega(x)).$$

This proves that the open facet Ω_F is the relative interior of the quasi-polyhedral face $\Omega(x)$.

Although the definition of a quasi-polyhedral face appears to be purely geometric, we show in our next result that a topological characterization is possible. This is the main result of this section and is employed in the following section to relate the analytic behavior of sequences in Ω to their geometric behavior.

THEOREM 2.8. *Let Ω_F be a nonempty face of the convex set Ω. Then Ω_F is a nonempty quasi-polyhedral face if and only if $\Omega_F + N(\Omega_F)$ has an interior. In either case,*

$$\text{int}\{\Omega_F + N(\Omega_F)\} = \text{ri}(\Omega_F) + \text{ri}(N(\Omega_F)).$$

Proof. Assume that $\Omega_F + N(\Omega_F)$ has an interior. Since (2.4) holds, we only need to show that

$$x + \text{lin}\{T(x)\} \subset \text{aff}(\Omega_F)$$

for any $x \in \text{ri}(\Omega_F)$. Since $(\text{aff}(\Omega_F) - x) + \text{span}\{N(\Omega_F)\}$ is a subspace with a nonempty interior,

$$R^n = (\text{aff}(\Omega_F) - x) + \text{span}\{N(\Omega_F)\}.$$

Now choose $z \in \text{lin}\{T(x)\}$. Then $z = x_1 + x_2$ where $x_1 \in \text{aff}(\Omega_F) - x$ and $x_2 \in \text{span}\{N(\Omega_F)\}$. We claim that $x_2 \in \text{span}\{N(\Omega_F)\}^\perp$, and thus that $x_2 = 0$. Note that (2.4) guarantees that $x_1 \in \text{lin}\{T(x)\}$, and since $z \in \text{lin}\{T(x)\}$ also, we must have $x_2 = z - x_1$ in $\text{lin}\{T(x)\}$. Lemma 2.7 then yields that $x_2 \in \text{span}\{N(\Omega_F)\}^\perp$ because $\text{span}\{S\}^\perp = S^\perp$ for any set S. We conclude that $x_2 = 0$, and thus z belongs to $\text{aff}(\Omega_F) - x$. Hence, $x + z \in \text{aff}(\Omega_F)$ as desired.

Now assume that Ω_F is a quasi-polyhedral face, and choose z in the set $\text{ri}(\Omega_F) + \text{ri}(N(\Omega_F))$. Then $z = x_1 + x_2$ where $x_1 \in \text{ri}(\Omega_F)$ and $x_2 \in \text{ri}(N(\Omega_F))$. We can also choose $\varepsilon > 0$ so that if B is the unit ball then

$$(x_1 + \varepsilon B) \cap \text{aff}(\Omega_F) \subset \Omega_F \tag{2.9}$$

and

$$(x_2 + \varepsilon B) \cap \text{aff}(N(\Omega_F)) \subset N(\Omega_F). \tag{2.10}$$

We now show that $z + \nu$ belongs to $\Omega_F + N(\Omega_F)$ for any $\nu \in R^n$ with $\|\nu\| \leqq \varepsilon$. Since Ω_F is quasi-polyhedral, Lemma 2.7 implies that

$$\text{aff}(\Omega_F) - x_1 = \text{lin}\{T(\Omega_F)\} = N(\Omega_F)^\perp.$$

Thus $\text{aff}(\Omega_F)-x_1$ is the orthogonal complement of the subspace $\text{aff}(N(\Omega_F))$, and therefore we can decompose ν as $\nu_1+\nu_2$ where $(\nu_1, \nu_2)=0$ and

$$\nu_1 \in \text{aff}(\Omega_F)-x_1, \qquad \nu_2 \in \text{aff}(N(\Omega_F)).$$

Clearly, $\|\nu_1\| \leqq \varepsilon$ and $x_1+\nu_1$ belongs to $\text{aff}(\Omega_F)$. Hence (2.9) implies that $x_1+\nu_1$ is in Ω_F. Similarly, $\|\nu_2\| \leqq \varepsilon$ and $x_2+\nu_2$ belongs to $\text{aff}(N(\Omega_F))$. Hence (2.10) implies that $x_2+\nu_2$ is in $N(\Omega_F)$. This shows that $z+\nu$ belongs to $\Omega_F+N(\Omega_F)$, and thus

$$\text{ri}(\Omega_F)+\text{ri}(N(\Omega_F)) \subset \text{int}\{\Omega_F+N(\Omega_F)\}.$$

The reverse inclusion is established by noting that

$$\text{int}\{\Omega_F+N(\Omega_F)\}=\text{ri}\{\Omega_F+N(\Omega_F)\}=\text{ri}(\Omega_F)+\text{ri}(N(\Omega_F)).$$

The first equality holds because the definition of the relative interior shows that if the interior of a convex set is not empty then the interior agrees with the relative interior. The second equality holds because the relative interior of the sum of convex sets is the sum of the relative interiors. This result is classical. See, for example, Corollary 6.6.2 of Rockafellar [1970]. □

3. Nondegeneracy and faces. Let $f: R^n \to R$ be a continuously differentiable mapping over a nonempty closed convex set $\Omega \subset R^n$, and consider a sequence $\{x_k\}$ which converges to a point $x^* \in \Omega$ that satisfies the first order necessary conditions for optimality for problem (1.1). Theorem 2.2 guarantees that $x^* \in \text{ri}(\Omega_F)$ for some face Ω_F of Ω. In this section we assume that Ω_F is quasi-polyhedral, and study conditions which guarantee that $x_k \in \text{ri}(\Omega_F)$ for all k sufficiently large.

Any point x^* which satisfies the first order necessary conditions for optimality of problem (1.1) is a *stationary point* for problem (1.1). The standard first order conditions are that

$$(\nabla f(x^*), x-x^*) \geqq 0, \qquad x \in \Omega.$$

An equivalent characterization in terms of the normal cone is that x^* is a stationary point if

$$-\nabla f(x^*) \in N(x^*). \tag{3.1}$$

If Ω is a convex set of the form

$$\Omega=\{x \in R^n : c_j(x) \geqq 0, j=1, \cdots, m\}, \tag{3.2}$$

for some differentiable functions $c_j: R^n \to R$, then stationary points are related to Kuhn-Tucker points if the constraint functions satisfy a constraint qualification. We shall use the Guignard [1969] constraint qualification which requires that

$$T(x)=\{\nu \in R^n : (\nu, \nabla c_i(x)) \geqq 0, j \in A(x)\}.$$

Gould and Tolle [1971] proved that the Guignard constraint qualification is the weakest condition that guarantees that a stationary point is a Kuhn-Tucker point. Note that if the constraint functions defining Ω are affine, then the Guignard constraint qualification is satisfied at every point in Ω. Also note that the Guignard constraint qualification implies that $N(x)$ is a *polyhedral cone*, that is, a cone of the form

$$K=\left\{\nu : \nu=\sum_{i=1}^{m} \lambda_i \nu_i, \lambda_i \geqq 0, \nu_i \in R^n\right\}. \tag{3.3}$$

In fact, the Farkas lemma implies that if the Guignard constraint qualification holds at $x \in \Omega$ then

$$N(x) = \left\{ \nu \in R^n : \nu = -\sum_{i \in A(x)} \lambda_i \nabla c_i(x),\ \lambda_i \geqq 0 \right\}.$$

Thus the normal cone is generated by the negatives of the active constraint gradients.

DEFINITION 3.1. A stationary point x^* of problem (1.1) is nondegenerate if

$$-\nabla f(x^*) \in \operatorname{ri}(N(x^*)).$$

Dunn [1987] introduced this definition of nondegeneracy and used it to show that the gradient projection method identifies the optimal face in a finite number of iterations. We shall show that this result can be extended considerably.

We claim that Definition 3.1 is a generalization of the strict complementarity condition. The classical definition of strict complementarity only applies to Kuhn–Tucker pairs when Ω is defined by a system of equations and inequalities. If the multipliers are not unique, then the active constraint normals are linearly dependent, and thus there is a set of Lagrange multipliers which fails the strict complementarity condition. We show this as follows. If

$$\nabla f(x) = \sum_{i \in A(x)} \lambda_i \nabla c_i(x), \qquad \lambda_i \geqq 0,$$

but the active constraint normals $\nabla c_i(x)$ are linearly dependent, then there are constants μ_i, not all zero, such that

$$\nabla f(x) = \sum_{i \in A(x)} (\lambda_i + \alpha \mu_i) \nabla c_i(x)$$

for any α. We can choose α so that $\lambda_i + \alpha\mu_i \geqq 0$, but with $\lambda_i + \alpha\mu_i = 0$ for at least one $i \in A(x)$. Thus $\lambda_i + \alpha\mu_i$ is a set of multipliers which fails the strict complementarity condition.

We now show that if Ω satisfies the Guignard constraint qualification then Definition 3.1 only requires the existence of some set of strictly complementary Lagrange multipliers. Since the Guignard constraint qualification implies that the normal cone is generated by the negative of the active constraint normals, we justify this claim by assuming that $N(x)$ is a polyhedral cone.

LEMMA 3.2. *If K is the polyhedral cone* (3.3) *then $\nu \in \operatorname{ri}(K)$ if and only if $\nu \in \operatorname{span}\{\nu_i : 1 \leqq i \leqq m\}$ with coefficients $\lambda_i > 0$ for $i = 1, \cdots, m$.*

Proof. Assume that $\nu \in \operatorname{span}\{\nu_i : 1 \leqq i \leqq m\}$ with coefficients $\lambda_i > 0$ for $i = 1, \cdots, m$. Note that if K is the polyhedral cone (3.3) then

$$\operatorname{aff}(K) = \operatorname{span}\{\nu_i : i = 1, \cdots, m\},$$

and let I be such that $\{\nu_i : i \in I\}$ is a basis for $\operatorname{aff}(K)$. The linear independence of $\{\nu_i : i \in I\}$ shows that there is an $\varepsilon > 0$ such that if $w \in \operatorname{aff}(K)$ and $\|w\| \leqq \varepsilon$, then

$$w = \sum_{i \in I} \mu_i \nu_i, \qquad |\mu_i| < \min\{\lambda_i : i = 1, \cdots, m\}.$$

Hence, $v + w$ is in K, and thus $v \in \operatorname{ri}(K)$. Conversely, assume that $v \in \operatorname{ri}(K)$, and choose any v_0 in $\operatorname{span}\{v_i : 1 \leqq i \leqq m\}$ with coefficients $\lambda_i > 0$ for $i = 1, \cdots, m$. Since $v \in \operatorname{ri}(K)$, there is an $\alpha > 1$ such that $\alpha\nu + (1-\alpha)\nu_0$ belongs to K. Hence,

$$\nu \in \left(\frac{1}{\alpha} K + \left(1 - \frac{1}{\alpha}\right) \nu_0 \right),$$

and thus $\nu \in \operatorname{span}\{\nu_i : 1 \leqq i \leqq m\}$ with coefficients $\lambda_i > 0$ for $i = 1, \cdots, m$. □

Lemma 3.2 shows that if Ω is a convex set of the form (3.2), and if the Guignard constraint qualification holds at x^*, then x^* is a nondegenerate Kuhn–Tucker point if and only if

$$\nabla f(x^*) = \sum_{i \in A(x^*)} \lambda_i \nabla c_i(x^*), \qquad \lambda_i > 0.$$

In other words, x^* is a nondegenerate Kuhn–Tucker point if and only if there is a set of strictly complementary Lagrange multipliers.

Definition 3.1 applies even in those cases where $N(x^*)$ fails to be polyhedral. For example, if Ω is the cone defined by

$$\Omega = \{(\xi_1, \xi_2, \xi_3): \xi_3 \geqq 0, \xi_3^2 \geqq \xi_1^2 + \xi_2^2\},$$

then $N(0)$ is not polyhedral, but it is clear that Definition 3.1 makes sense. Note that the constraint functions which define Ω do not satisfy the Guignard constraint qualification at the origin.

The characterization of algorithms that generate sequences $\{x_k\}$ such that $x_k \in \mathrm{ri}\,(\Omega_F)$ for all k sufficiently large is in terms of the *projected gradient* $\nabla_\Omega f$ which is defined by

$$\nabla_\Omega f(x) \equiv \operatorname{argmin}\{\|\nu + \nabla f(x)\|: \nu \in T(x)\}.$$

Since $T(x)$ is a nonempty closed convex set, this defines $\nabla_\Omega f(x)$ uniquely. Note that $\nabla_\Omega f(x)$ is the projection of $-\nabla f(x)$ into $T(x)$, that is,

$$\nabla_\Omega f(x) = P_{T(x)}[-\nabla f(x)].$$

This definition of the projected gradient was used by Calamai and Moré [1987] in their work on the projected gradient. They showed, in particular, that if $\{x_k\}$ is the sequence generated by the gradient projection algorithm, then $\{\nabla_\Omega f(x_k)\}$ converges to zero. In the next section we show that sequential quadratic programming algorithms also force $\{\nabla_\Omega f(x_k)\}$ to zero.

The projected gradient shares many properties with the standard gradient. For example, $\nabla_\Omega f(x^*) = 0$ if and only if x^* is a stationary point for problem (1.1). In general $\nabla_\Omega f$ is not continuous because it depends on the tangent cone. However, Calamai and Moré [1987] prove that $\|\nabla_\Omega f(\cdot)\|$ is lower semicontinuous, and Theorem 2.3 shows that $\nabla_\Omega f$ is continuous on the relative interior of any face of Ω.

The proof of the characterization result relies on the following technical lemma.

LEMMA 3.3. *Assume that Ω_F is a quasi-polyhedral face of the convex set Ω with $x^* \in \mathrm{ri}\,(\Omega_F)$ and $d^* \in \mathrm{ri}\,(N(\Omega_F))$. If $x_k \in \Omega$ and $d_k \in N(x_k)$, and the sequences $\{x_k\}$ and $\{d_k\}$ converge to x^* and d^*, respectively, then $x_k \in \mathrm{ri}\,(\Omega_F)$ for all k sufficiently large.*

Proof. Theorem 2.8 guarantees that

$$x^* + d^* \in \mathrm{int}\,\{\Omega_F + N(\Omega_F)\},$$

and since $\{x_k + d_k\}$ converges to $x^* + d^*$,

$$x_k + d_k \in \mathrm{int}\,\{\Omega_F + N(\Omega_F)\}.$$

We now claim that if P is the projection into Ω, then $P(u+v) = u$ for any $u \in \Omega$ and $v \in N(u)$. The validity of this observation is established by noting that $P(x)$ is characterized by the requirement that $x - P(x)$ be in the normal cone $N(P(x))$. Hence, Theorem 2.8 and this observation yield that

$$x_k = P(x_k + d_k) \in P(\mathrm{ri}\,(\Omega_F) + \mathrm{ri}\,(N(\Omega_F))) \tag{3.4}$$

for all k sufficiently large. Now note that if $u \in \text{ri}(\Omega_F)$ and $v \in \text{ri}(N(\Omega_F))$ then Theorem 2.3 implies that $v \in N(u)$. Hence, $P(u+v) = u \in \text{ri}(\Omega_F)$. In particular,

$$P(\text{ri}(\Omega_F) + \text{ri}(N(\Omega_F))) \subset \text{ri}(\Omega_F).$$

Thus (3.4) implies that $x_k \in \text{ri}(\Omega_F)$ as desired. □

The gradient and the projected gradient are related via the Moreau decomposition: If K is a closed convex cone then every $x \in R^n$ can be uniquely expressed as

$$x = P_K(x) + P_{K^\circ}(x).$$

This result can be found, for example, as Lemma 2.2 of Zarantonello [1971]. An immediate consequence of this decomposition is that

$$\|\nabla_\Omega f(x)\| = \text{dist}(-\nabla f(x), N(x)),$$

where

$$\text{dist}(y, \Gamma) \equiv \inf\{\|y-x\| : x \in \Gamma\}$$

is the distance function for the set Γ. We now show that the main result of this section is a consequence of the Moreau decomposition and Lemma 3.3.

THEOREM 3.4. *Let $f: R^n \to R$ be continuously differentiable over a nonempty closed convex set Ω, and assume that $\{x_k\}$ is a sequence in Ω which converges to a nondegenerate stationary point z^*. If Ω_F is a quasi-polyhedral face of Ω with $x^* \in \text{ri}(\Omega_F)$, then $x_k \in \text{ri}(\Omega_F)$ for all k sufficiently large if and only if $\{\nabla_\Omega f(x_k)\}$ converges to zero.*

Proof. If $x_k \in \text{ri}(\Omega_F)$ for all k sufficiently large, then Theorem 2.3 shows that

$$\nabla_\Omega f(x_k) = P_{T(x^*)}[-\nabla f(x_k)].$$

Since ∇f is continuous, and since x^* is a stationary point of problem (1.1), we obtain that $\{\nabla_\Omega f(x_k)\}$ converges to zero. For the converse define

$$d_k = P_{N(x_k)}[-\nabla f(x_k)],$$

and note that the Moreau decomposition of $-\nabla f(x_k)$ yields

$$-\nabla f(x_k) = \nabla_\Omega f(x_k) + d_k.$$

Since $\{\nabla_\Omega f(x_k)\}$ converges to zero, and since x^* is nondegenerate,

$$\lim_{k\to\infty} d_k = d^* \equiv -\nabla f(x^*) \in \text{ri}(N(\Omega_F)).$$

Thus, Lemma 3.3 shows that $x_k \in \text{ri}(\Omega_F)$ for all k sufficiently large. □

We now show that if x^* is a nondegenerate stationary point and if $N(x^*)$ has a nonempty interior, then finite termination is obtained.

COROLLARY 3.5. *Let the hypothesis of Theorem 3.4 hold. If $N(x^*)$ has a nonempty interior and $\{\nabla_\Omega f(x_k)\}$ converges to zero, then $x_k = x^*$ for all k sufficiently large.*

Proof. The result follows from Theorem 3.4 once we show that $\{x^*\}$ is a quasi-polyhedral face. If $\Omega(x^*)$ is the face of Ω with $x^* \in \text{ri}(\Omega(x^*))$ then $\Omega(x^*) + N(x^*)$ has an interior, and thus Theorem 2.8 shows that $\Omega(x^*)$ is quasi-polyhedral. Moreover, since $\text{int}(N(x^*))$ is not empty, the linearity of $T(x^*)$ is $\{0\}$. Thus $\{x^*\} = \Omega(x^*)$ is a quasi-polyhedral face. □

The proof of this result shows that if $N(x^*)$ has a nonempty interior then $\{x^*\}$ is a quasi-polyhedral face of Ω. The converse of this statement is easily established. We also claim that if $N(x^*)$ has a nonempty interior then x^* is an extreme point of Ω. We prove this by assuming that $x^* = \lambda x_1 + (1-\lambda)x_2$ for some $0 < \lambda < 1$ and x_1, x_2 in

Ω. Then span $\{x_2 - x_1\}$ is in $T(x^*)$, and thus $x_1 = x_2$. Note that the converse fails because an extreme point of a general convex set may have $N(x^*)$ with a nonempty interior. However, if Ω is a polyhedron then any extreme point x^* has a normal cone $N(x^*)$ with a nonempty interior.

If Ω is a polyhedron then any x^* in Ω is in the relative interior of a quasi-polyhedral face. Theorem 3.4 simplifies in this situation, and we can express our results in terms of the active constraints.

COROLLARY 3.6. *Let $f: R^n \to R$ be continuously differentiable on the polyhedron Ω, and assume that $\{x_k\}$ is a sequence in Ω which converges to a nondegenerate stationary point x^*. Then $A(x_k) = A(x^*)$ for all k sufficiently large if and only if $\{\nabla_\Omega f(x_k)\}$ converges to zero.*

Proof. We only need to show that $A(x)$ is independent of $x \in \text{ri}(\Omega_F)$. If Ω is the polyhedron (2.5) and x_1 and x_2 are in ri (Ω_F), then by definition, $x_1 - x_2$ is in lin $\{T(x_1)\}$. Thus (2.6) implies that if $i \in A(x_1)$ then $(c_i, x_1 - x_2) = 0$. Hence, $A(x_1)$ is a subset of $A(x_2)$. Since the argument is symmetric in x_1 and x_2, the result holds. □

The characterization result of Corollary 3.6 extends and generalizes the result of Calamai and Moré [1987] which states that if $\{x_k\}$ is any sequence in Ω which converges to a point x^*, which is nondegenerate stationary in the standard sense, then $A(x_k) = A(x^*)$ for all k sufficiently large provided $\{\nabla_\Omega f(x_k)\}$ converges to zero.

4. Algorithms. We have established that if the stationary point is nondegenerate and lies in the relative interior of a quasi-polyhedral face, then any algorithm that forces the projected gradient $\nabla_\Omega f$ to zero attains the optimal face in a finite number of iterations. In particular, if Ω is a polyhedron, then any such algorithm identifies the optimal active constraints in a finite number of iterations, and if the Kuhn–Tucker point x^* is an extreme point of Ω, then x^* is attained in a finite number of iterations. In this section we show how these results apply to the sequential quadratic programming algorithm and to the gradient projection algorithm.

We first examine the sequential quadratic programming algorithm. Let $f: R^n \to R$ be a continuously differentiable mapping over a nonempty closed set $\Omega \subset R^n$, and consider the subproblem

$$\min \{q_k(w): x_k + w \in \Omega\} \tag{4.1}$$

where $x_k \in \Omega$ and $q_k: R^n \to R$ is the quadratic

$$q_k(w) = (\nabla f(x_k), w) + \tfrac{1}{2}(w, B_k w),$$

for some symmetric matrix $B_k \in R^{n \times n}$. If p_k is a solution of subproblem (4.1) then the next iterate is defined by setting $x_{k+1} = x_k + \alpha_k p_k$ where $\alpha_k > 0$ is the step.

For a general convex set Ω and $B_k \equiv 0$, the algorithm outlined above is the conditional gradient algorithm, while if $B_k = \nabla^2 f(x_k)$, this algorithm is Newton's method for the constrained problem (1.1). If Ω is a polyhedron then subproblem (4.1) reduces to the standard sequential quadratic programming algorithm for linearly constrained problems. Indeed, $x_k + w$ belongs to the polyhedron (2.5) if and only if $(c_j, x_k + w) \geqq \delta_j$ for $1 \leqq j \leqq m$.

It is worthwhile noting that an algorithm based on (4.1) can be implemented efficiently for certain kinds of nonpolyhedral convex sets. For example, if

$$\Omega = \{x \in R^n: (x, Dx) \leqq \delta\}$$

for some positive definite matrix $D \in R^{n \times n}$ and scalar $\delta > 0$, then the algorithm developed by Moré and Sorensen [1983] determines the global minimizer of (4.1) for a general matrix B_k.

We do not elaborate on the choice of B_k and α_k since it is not our purpose to establish the convergence of this method. Convergence results for the conditional gradient method and for Newton's method can be found, for example, in the works of Daniel [1971], Pshenichny and Danilin [1978], and Dunn [1980]. For linearly constrained problems, Garcia-Palomares [1975] established convergence of the sequential quadratic programming algorithm. Other convergence results can be obtained by specializing the results for nonlinear programming problems. See, for example, Burke and Han [1986], Powell [1983], and Han [1977].

In our analysis we restrict our attention to those cases in which $\{B_k\}$ is bounded and the above algorithm returns a convergent sequence $\{x_k\}$ whose limit x^* is a stationary point for problem (1.1). We also assume that $\{p_k\}$ converges to zero. This assumption is satisfied for the standard choices of α_k if, for example, B_k is uniformly positive definite for all k sufficiently large.

THEOREM 4.1. *Let $f: R^n \to R$ be continuously differentiable over a nonempty closed convex set Ω, and assume that $\{x_k\}$ is a bounded sequence in Ω and that p_k is a stationary point of* (4.1). *If $\{B_k\}$ is bounded and $\{p_k\}$ converges to zero, then $\{\nabla_\Omega f(x_k+p_k)\}$ converges to zero.*

Proof. Let $N(\Gamma, x)$ be the normal cone of the convex set Γ at x. Since p_k is a stationary point of subproblem (4.1), the first order conditions for this subproblem imply that

$$-\nabla q_k(p_k) \in N(\Omega - x_k, p_k) = N(\Omega, x_k + p_k). \tag{4.2}$$

The relationship $N(\Omega - x_k, p_k) = N(\Omega, x_k + p_k)$ follows from the definition of a normal cone. We now claim that

$$\|\nabla_\Omega f(x_k + p_k)\| \le \|\nabla f(x_k + p_k) - \nabla q_k(p_k)\|. \tag{4.3}$$

This claim is verified by noting that the Moreau decomposition implies that for any $x \in \Omega$ and any $\nu \in N(x)$,

$$\|\nabla_\Omega f(x)\| = \|\nabla f(x) + P_{N(x)}[-\nabla f(x)]\| \leqq \|\nabla f(x) + \nu\|.$$

Since $x_k + p_k$ is in Ω, inequality (4.3) is a direct consequence of condition (4.2). Thus

$$\|\nabla_\Omega f(x_k + p_k)\| \leqq \|\nabla f(x_k + p_k) - \nabla f(x_k) - B_k p_k\|,$$

and the result follows by appealing to the uniform continuity of ∇f on bounded sets, the convergence of $\{p_k\}$ to zero and the boundedness of $\{B_k\}$. □

If Ω is a polyhedron, Theorem 4.1 and Corollary 3.6 imply that if the sequence $\{x_k\}$ generated by the sequential quadratic programming method converges to a nondegenerate stationary point x^*, then $A(x_k + p_k) = A(x^*)$ for all k sufficiently large. Note that there is no need to assume linear independence of the active constraint normals; the strict complementarity condition in the sense of Definition 3.1 is all that is needed.

If Ω is a general convex set, Theorem 3.4 and 4.1 show that if x^* is in the relative interior of a quasi-polyhedral face Ω_F then $x_k + p_k$ belongs to ri (Ω_F) for all k sufficiently large. Thus, the subproblem (4.1) identifies the optimal face in a finite number of iterations.

We now turn to the application of the results of § 3 to the gradient projection method (1.3). The step α_k is defined is terms of positive constants γ_1, γ_2, and γ_3, and constants μ and η in $(0, 1)$ with $\mu < \eta$. Define

$$\bar{x}_{k+1} = P(x_k - \bar{\alpha}_k \nabla f(x_k))$$

where $\bar{\alpha}_k > 0$ satisfies

$$f(\bar{x}_{k+1}) > f(x_k) + \eta \langle \nabla f(x_k), \bar{x}_{k+1} - x_k \rangle.$$

We assume that the step α_k satisfies

$$f(x_{k+1}) \leqq f(x_k) + \mu \langle \nabla f(x_k), x_{k+1} - x_k \rangle, \tag{4.4}$$

and

$$\gamma_3 \geqq \alpha_k \geqq \gamma_1 \quad \text{or} \quad \gamma_3 \geqq \alpha_k \geqq \gamma_2 \bar{\alpha}_k. \tag{4.5}$$

Condition (4.4) on α_k forces a sufficient decrease of the function while condition (4.5) guarantees that α_k is not too small. Under these conditions Calamai and Moré [1987] established the following convergence result for the gradient projection method.

THEOREM 4.2. *Let $f: R^n \to R$ be continuously differentiable on Ω, and let $\{x_k\}$ be the sequence generated by the gradient projection method. If f is bounded below on Ω and ∇f is uniformly continuous on Ω then*

$$\lim_{k\to\infty} \|\nabla_\Omega f(x_k)\| = 0.$$

Theorems 4.2 and Corollary 3.6 imply that if Ω is a polyhedron, and if the sequence $\{x_k\}$ generated by the gradient projection method converges to a nondegenerate stationary point x^*, then $A(x_k) = A(x^*)$ for all k sufficiently large. This result had been obtained by Calamai and Moré [1987] under the additional assumption that the active constraint normals are linearly independent.

Theorem 4.2 can also be combined with Theorem 3.4 to show that the gradient projection method forces $x_k \in \operatorname{ri}(\Omega_F)$ for all k sufficiently large. Dunn [1987] had obtained this result under the additional assumption that the steps $\{\alpha_k\}$ are bounded away from zero; we note that the steps are bounded away from zero if ∇f is Lipschitz continuous and α_k satisfies (4.4) and (4.5). We also note that the approach followed by Dunn to establish identification results is direct, and does not show that $\{\nabla_\Omega f(x_k)\}$ converges to zero.

There are other results that show that the optimal face is identified in a finite number of iterations, but these results require standard linear independence and strict complementarity conditions at the stationary point. For example, this is the case for the generalization of the gradient projection algorithm considered by Gawande and Dunn [1988].

5. Subdifferentials and projected gradients. Let $f: R^n \to R$ be a continuously differentiable mapping over the closed convex set $\Omega \subset R^n$, and let $P: R^n \to \Omega$ be the projection into Ω. The differentiable constrained problem (1.1) can be transformed into the nondifferentiable unconstrained problem

$$\min \{\phi(x) : x \in R^n\},$$

where the composite mapping $\phi: R^n \to R$ is defined by

$$\phi(x) = f(P(x)). \tag{5.1}$$

The mapping ϕ is locally Lipschitz on R^n, and thus has a Clarke subdifferential $\partial\phi(x)$ at each point in R^n. In this section we establish an interesting relationship between $\nabla_\Omega f$ and $\partial\phi$. We show that if $\{\|\nabla_\Omega f(x_k)\|\}$ converges to zero, then $\{\operatorname{dist}(0, \partial\phi(x_k))\}$ also converges to zero.

This result is remarkable because in nondifferentiable optimization $\{\text{dist}(0, \partial\phi(x_k))\}$ almost always stays bounded away from zero. We have shown that projected gradient and sequential quadratic programming algorithms force $\{\|\nabla_\Omega f(x_k)\|\}$ to zero, so this result shows that $\{\text{dist}(0, \partial\phi(x_k))\}$ converges to zero for these algorithms.

We begin our analysis by obtaining an expression for the directional derivative of ϕ. Lemma 4.6 of Zarantonello [1971] shows that if $x \in \Omega$ and $\nu \in R^n$ then

$$\lim_{\alpha \to 0^+} \frac{P(x+\alpha\nu) - P(x)}{\alpha} = P_{T(x)}(\nu).$$

Another proof of this result can be found in the paper of McCormick and Tapia [1972]. From this result we immediately obtain that the directional derivative of ϕ is

$$\phi'(x; \nu) = (\nabla f(x), P_{T(x)}(\nu)). \tag{5.2}$$

We now relate the directional derivative of ϕ to the projected gradient of f. We claim that

$$\min\{\phi'(x; \nu) : \|\nu\| \leqq 1\} = \min\{(\nabla f(x), \nu) : \nu \in T(x), \|\nu\| \leqq 1\}.$$

The proof of this claim follows from (5.2) and the observation that projections are nonexpansive mappings so that

$$\|P_{T(x)}(\nu)\| = \|P_{T(x)}(\nu) - P_{T(x)}(0)\| \leqq \|\nu\|.$$

Since Lemma 3.1 of Calamai and Moré [1987] guarantees that

$$\min\{(\nabla f(x), \nu) : \nu \in T(x), \|\nu\| \leqq 1\} = -\|\nabla_\Omega f(x)\|,$$

we obtain that

$$\min\{\phi'(x; \nu) : \|\nu\| \leqq 1\} = -\|\nabla_\Omega f(x)\|. \tag{5.3}$$

The validity of (5.3) depends on the relationship (5.1) between ϕ and f. The following result, which can be found in Burke [1983], only assumes that $\phi : R^n \to R$ is a locally Lipschitz function.

LEMMA 5.1. *Let $\phi : R^n \to R$ be a locally Lipschitz function of R^n, and let $\phi^\circ(x; \cdot)$ be the generalized directional derivative at $x \in R^n$. Then*

$$\min\{\phi^\circ(x; \nu) : \|\nu\| \leqq 1\} = -\text{dist}(0, \partial\phi(x)). \tag{5.4}$$

Proof. The proof uses two standard results about the subdifferential. They are that

$$\phi^\circ(x; \nu) = \max\{(\xi, \nu) : \xi \in \partial\phi(x)\},$$

and that $\partial\phi(x)$ is a nonempty compact convex set. Clarke [1983] proves these results. We establish (5.4) by showing that for any closed convex set Γ

$$\min\{\max\{(\xi, \nu) : \xi \in \Gamma\} : \|\nu\| \leqq 1\} = -\text{dist}(0, \Gamma). \tag{5.5}$$

Choose $\xi^* = P_\Gamma(0)$. For any $\|\nu\| \leqq 1$ we have $(\xi^*, \nu) \geqq -\|\xi^*\|$, and hence

$$\max\{(\xi, \nu) : \xi \in \Gamma\} \geqq -\|\xi^*\| = -\text{dist}(0, \Gamma).$$

Since this holds for all $\|\nu\| \leqq 1$,

$$\min\{\max\{(\xi, \nu) : \xi \in \Gamma\} : \|\nu\| \leqq 1\} \geqq -\text{dist}(0, \Gamma).$$

A similar argument yields the reverse inequality. Choose $\xi^* = P_\Gamma(0)$. Then $(\xi^*, \xi - \xi^*) \geqq 0$ for all $\xi \in \Gamma$, and if $\nu^* = -\xi^*/\|\xi^*\|$ when $\xi^* \neq 0$ and $\nu^* = 0$ otherwise, then

$$\max\{(\xi, \nu^*) : \xi \in \Gamma\} \leqq -\|\xi^*\| = -\text{dist}(0, \Gamma).$$

Since $\|\nu^*\| \leqq 1$, this yields the reverse inequality and establishes (5.5). □

We now have all the ingredients needed to prove the main result of this section.

THEOREM 5.2. *Let $f: R^n \to R$ be continuously differentiable over a nonempty closed convex set Ω, and define ϕ by* (5.1). *Then*

$$\text{dist}\,(0, \partial\phi(x)) \leqq \|\nabla_\Omega f(x)\|. \tag{5.6}$$

Proof. Since $\phi'(x; \nu) \leqq \phi^\circ(x; \nu)$, the proof is an immediate consequence of (5.3) and (5.4). □

Inequality (5.6) can be strict. For example, if Ω is the positive orthant R^n_+, then $\nabla_\Omega f(0) = 0$ if and only if $\nabla f(0) \in R^n_+$ but $0 \in \partial\phi(0)$ for any function f. The last part of this claim follows from (5.4) if we show that $\phi^\circ(0; \nu) \geqq 0$. Recall that

$$\phi^\circ(x; \nu) \equiv \limsup_{y \to x, \tau \to 0^+} \frac{\phi(y + \tau\nu) - \phi(y)}{\tau}.$$

If we consider sequences y_k and τ_k such that $P(y_k + \tau_k \nu) = P(y_k) = P(0)$, it is clear that $\phi^\circ(0; \nu) \geqq 0$.

REFERENCES

D. P. BERTSEKAS [1976], *On the Goldstein-Levitin-Polyak projection method*, IEEE Trans. Automat. Control, 21, pp. 174–184.

J. V. BURKE [1983], *Methods for solving generalized systems of inequalities with application to nonlinear programming*, Ph.D. dissertation, University of Illinois, Urbana, IL.

J. V. BURKE AND S. P. HAN [1988], *Robust sequential quadratic programming methods*, to appear in Math. Prog.

P. H. CALAMAI AND J. J. MORÉ [1987], *Projected gradient methods for linearly constrained problems*, Math. Prog., 39, pp. 93–116.

F. H. CLARKE [1983], *Optimization and Nonsmooth Analysis*, John Wiley, New York.

J. W. DANIEL [1971], *The Approximate Minimization of Functionals*, Prentice-Hall, Englewood Cliffs, NJ.

J. C. DUNN [1980], *Newton's method and the Goldstein step-length rule for constrained minimization problems*, SIAM J. Control Optim., 18, pp. 659–674.

—— [1981], *Global and asymptotic convergence rate estimates for a class of projected gradient processes*, SIAM J. Control Optim., 19, pp. 368–400.

—— [1987], *On the convergence of projected gradient processes to singular critical points*, J. Optim. Theory Appl., 56, pp. 203–216.

U. M. GARCIA-PALOMARES [1975], *Superlinearly convergent algorithms for linearly constrained optimization*, in Nonlinear Programming 2, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York.

M. GAWANDE AND J. C. DUNN [1988], *Variable metric gradient projection processes in convex feasible sets defined by nonlinear inequalities*, Appl. Math. Optim., 17, pp. 103–119.

F. J. GOULD AND J. W. TOLLE [1971], *A necessary and sufficient qualification for constrained optimization*, SIAM J. Appl. Math., 20, pp. 164–172.

M. GUIGNARD [1969], *Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space*, SIAM J. Control, 7, pp. 232–241.

S. P. HAN [1977], *A globally convergent method for nonlinear programming*, J. Optim. Theory Appl., 22, pp. 297–309.

G. P. MCCORMICK AND R. A. TAPIA [1972], *The gradient projection method under mild differentiability conditions*, SIAM J. Control, 10, pp. 93–98.

J. J. MORÉ AND D. C. SORENSEN [1983], *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4, pp. 553–572.

M. J. D. POWELL [1983], *Variable metric methods for constrained optimization*, in Mathematical Programming Bonn 1982—The State of the Art, A. Bachem, M. Grötschel, and B. Korte, eds., Springer-Verlag, Berlin, New York.

B. N. PSHENICHNY AND YU. M. DANILIN [1978], *Numerical Methods in Extremal Problems*, MIR, Moscow.

R. T. ROCKAFELLAR [1970], *Convex Analysis*, Princeton University Press, Princeton, NJ.

E. H. ZARANTONELLO [1971], *Projections on convex sets in Hilbert space and spectral theory*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, New York.