ARGONNE NATIONAL LABORATORY
9700 South Cass Avenue
Argonne, Illinois 60439

EXPOSING CONSTRAINTS

James V. Burke* **and Jorge J. Moré**

Mathematics and Computer Science Division

Preprint MCS-P308-0592

June 1992

Work 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.

* Permanent address: Department of Mathematics, University of Washington, Seattle, Washington 98195. Work supported in part by the National Science Foundation grant DMS-9102059.

ABSTRACT

The development of algorithms and software for the solution of large-scale optimization problems has been the main motivation behind the research on the identification properties of optimization algorithms. The aim of an identification result for a linearly constrained problem is to show that if the sequence generated by an optimization algorithm converges to a stationary point, then there is a nontrivial face F of the feasible set such that after a finite number of iterations, the iterates enter and remain in the face F. This paper develops the identification properties of linearly constrained optimization algorithms without any non-degeneracy or linear independence assumptions. The main result shows that the projected gradient converges to zero if and only if the iterates enter and remain in the face exposed by the negative gradient. This result generalizes results of Burke and Moré obtained for nondegenerate cases.

EXPOSING CONSTRAINTS

James V. Burke* and Jorge J. Moré

1 Introduction

The development of algorithms and software for the solution of large-scale constrained optimization problems,

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

has been the main motivation behind the research on the identification properties of optimization algorithms. Much of this research has been done under the assumption that the set Ω is a general convex set and that the stationary points of the optimization problem (1.1) are nondegenerate. In this work we show that it is possible to develop the identification properties of linearly constrained optimization algorithms without any nondegeneracy or linear independence assumptions.

The aim of an identification result is to show that if $\{x_k\}$ is a sequence in Ω that converges to a stationary point x^*, then there is an index $k_0 > 0$ and a nontrivial face $F(x^*)$ of Ω with $x_k \in F(x^*)$ for all $k \geq k_0$. These results are of importance because they show that eventually the behavior of the algorithm is determined by the properties of f on the face $F(x^*)$. For recent work on the identification properties of optimization algorithms see, for example, Conn, Gould, and Toint [7]. Dunn [11], Wright [24], Burke [3], Burke, Moré, and Toraldo [5], Lescrenier [17], Wright [25], and Kelley and Sachs [16].

In linearly constrained problems an identification result can also be expressed in terms of the indices of the active constraints. For example, if Ω has the specific representation

$$\Omega = \{x \in \mathbb{R}^n : \langle c_j, x \rangle \geq \delta_j,\ 1 \leq j \leq m\}, \tag{1.2}$$

for some vectors $c_j \in \mathbb{R}^n$ and scalars δ_j, then all faces F of Ω are of the form

$$F = \{x \in \Omega : \langle c_j, x \rangle = \delta_j,\ j \in \mathcal{A}(F)\}$$

for some index set $\mathcal{A}(F)$. Thus, identification results can be given in terms of an index set that defines the face $F(x^*)$. In this paper we obtain results in terms of the face structure of Ω; results in terms of index sets of active constraints are direct consequences of these results.

An important consequence of the assumption that the stationary point x^* is nondegenerate is that $F(x^*)$ is the unique face of Ω that contains x^* in the relative interior. Thus, it

Work 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.

*Permanent address: Department of Mathematics, University of Washington, Seattle, Washington 98195. Work supported in part by the National Science Foundation grant DMS-9102059.

follows that $x_k \in \operatorname{ri}\{F(x^*)\}$ for all $k \geq k_0$, where $\operatorname{ri}\{\cdot\}$ denotes the relative interior. One of the main difficulties in extending identification results to the degenerate case is that there may be no face $F(x^*)$ such that $x_k \in \operatorname{ri}\{F(x^*)\}$ for all k sufficiently large. However, we show that there is a face $F(x^*)$ such that $x_k \in F(x^*)$ for all k sufficiently large.

The approach in this paper is reminiscent of the approach of Burke and Moré [4]. In particular, the approach depends on the concept of an exposed face, and on the facial geometry of a convex set Ω. These results are developed in Sections 2 and 3. The definition of an exposed face and several important properties of exposed faces are presented in Section 2, while Section 3 contains a key result on the existence of a partition of $\mathbb{R}^n$ based on the face structure of Ω.

The main result of this paper appears in Section 4. This result is a characterization of the identification properties of a sequence $\{x_k\}$ in terms of the projected gradient and the face $E[-\nabla f(x^*)]$ of Ω exposed by $-\nabla f(x^*)$. If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable on the polyhedral set Ω, and $\{x_k\}$ is a sequence in Ω that converges to a stationary point x^* of (1.1), then we show that

$$\lim_{k \to \infty} P_{T(x_k)}[-\nabla f(x_k)] = 0 \tag{1.3}$$

if and only if there is a $k_0 > 0$ such that

$$x_k \in E[-\nabla f(x^*)], \qquad k \geq k_0. \tag{1.4}$$

This result has important ramifications because optimization algorithms for linearly constrained problems tend to satisfy (1.3). This can be seen by noting that if the polyhedral set Ω has the representation (1.2), then the projected gradient is

$$P_{T(x)}[-\nabla f(x)] = -\nabla f(x) + \sum_{j \in \mathcal{A}(x)} \lambda_j c_j,$$

where $\mathcal{A}(x)$ is the set of active constraints at $x \in \Omega$, and λ_j is a nonnegative estimate (a precise definition is given in Section 6) of the Lagrange multiplier for the i-th constraint.

The identification property (1.4) can be expressed in terms of Lagrange multipliers at the solution because we show that $E[-\nabla f(x^*)]$ is defined in terms of the active set by

$$x \in E[-\nabla f(x^*)] \iff \{i \in \mathcal{A}(x^*) : \lambda_i^* > 0\} \subset \mathcal{A}(x), \quad x \in \Omega,$$

where λ_i^* is the Lagrange multiplier for the i-th constraint. This result depends on the choice of Lagrange multipliers.

The identification property (1.4) has an equivalent formulation that is clearly independent of the choice of Lagrange multipliers. Section 5 shows that it is possible to define a set of strictly binding constraints $\mathcal{B}_s^*$ at a stationary point x^* independent of the choice of Lagrange multipliers, and that

$$x \in E[-\nabla f(x^*)] \iff \mathcal{B}_s^* \subset \mathcal{A}(x), \quad x \in \Omega.$$

This leads, in particular, to a version of the results of Section 4 in terms of $\mathcal{B}_s^*$.

Section 4 also contains a discussion of the connection of these result to the identification results of Burke and Moré [4], and to the convergence results for the class of trust region methods for bound constrained optimization problems proposed by Conn, Gould, and Toint [7, 8]. The connection between these results and the class of trust region methods for general linearly constrained methods analyzed by Moré [18] and Burke, Moré, and Toraldo [5] will be reported in a later paper.

We end the paper with two applications of these identification results. Section 6 examines the influence of degeneracy on the standard second-order sufficiency conditions. In particular, we show that if x^* is degenerate, a satisfactory convergence analysis can be obtained if we assume that $\nabla^2 f(x^*)$ is positive definite on the affine hull of $E[-\nabla f(x^*)] - x^*$. Section 7 examines the implication of the identification results to the GPCG algorithm of Moré and Toraldo [19]. In particular, we show that the convergence condition of the GPCG algorithm is satisfied in a finite number of iterations even when the solution is degenerate.

2 Exposed Faces

The geometric approach to the identification properties of optimization algorithms requires an understanding of the face structure of a convex set Ω. In this section we provide some of the necessary results and background from convex analysis.

Recall that the *affine hull* $\mathrm{aff}\{\Omega\}$ of a convex set Ω in $\mathbb{R}^n$ is the smallest affine set that contains Ω, and the *relative interior* $\mathrm{ri}\,(\Omega)$ of Ω is the interior of Ω relative to $\mathrm{aff}\{\Omega\}$. In all cases we assume that Ω is not empty.

A nonempty subset F of a convex set Ω is a *face* of Ω if every convex subset of Ω whose relative interior meets F is contained in F. Thus, if x and y are in Ω and $\lambda x + (1-\lambda)y$ lies in F for some λ in $(0,1)$, then both x and y must belong to F. A basic result on the face structure of a convex set is that the relative interiors of the faces of Ω form a partition of Ω. For future reference, we state this result formally.

Theorem 2.1 *If $\mathcal{F}$ is the collection of all faces of the convex set Ω, then the collection*

$$\{\mathrm{ri}\,(F) : F \in \mathcal{F}\}$$

is a partition of Ω.

This result can be found, for example, in Rockafellar [22, Theorem 18.2]. Note that this result shows that every point $x \in \Omega$ can be associated with a unique face $F(x)$ of Ω such that $x \in \mathrm{ri}\,(F)$.

We are concerned with faces of Ω that are exposed by a given vector d in $\mathbb{R}^n$. A nonempty subset F of a convex set Ω is *exposed* by a vector $d \in \mathbb{R}^n$ if $F = E(d)$, where

$$E(d) = \arg\max\{\langle d, x\rangle : x \in \Omega\}$$

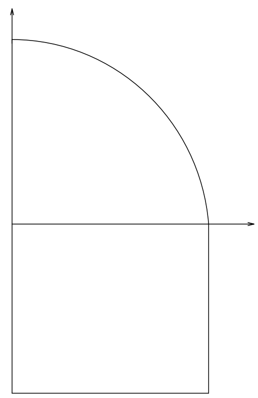

Figure 2.1: Convex set defined by (2.1)

for some inner product $\langle \cdot, \cdot \rangle$ of $\mathbb{R}^n$. A computation shows that $E(d)$ is a face whenever $E(d)$ is not empty.

Every face of a polyhedron is exposed (see, for example, Theorem 2.4.12 of Stoer and Witzgall [23]), but this is not the case for general convex sets. For example, in the convex set Ω defined by

$$\Omega = \left\{ (\xi_1, \xi_2) : -1 \le \xi_2 \le (1 - \xi_1^2)^{\frac{1}{2}},\ 0 \le \xi_1 \le 1 \right\}, \tag{2.1}$$

the point $(1, 0)$ is a face of Ω that is not exposed by any vector. This can be seen clearly in Figure 2.1.

The concept of an exposed face is closely related to the concept of a normal cone. For a convex set Ω, the normal cone at x in Ω is defined by

$$N(x) = \{ u \in \mathbb{R}^n : \langle u, y - x \rangle \le 0,\ y \in \Omega \}.$$

The tangent cone $T(x)$ is the dual of the normal cone. Thus $v \in T(x)$ if and only if $\langle v, u \rangle \le 0$ for all $u \in N(x)$. The tangent cone $T(x)$ can also be defined as the closure of all vectors $v \in \mathbb{R}^n$ such that $x + \alpha v \in \Omega$ for all $\alpha > 0$ sufficiently small.

Exposed faces and normal cones are related by the observation that $x \in E(d)$ if and only if $d \in N(x)$. We make use of this observation throughout this section.

The concept of an exposed face is also related to the standard first-order optimality conditions for problem (1.1). Recall that the standard first order conditions for a stationary point x^* of problem (1.1) are that

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

In terms of normal cones, this is equivalent to requiring that

$$-\nabla f(x^*) \in N(x^*).$$

The following result shows the connection between stationary points and exposed faces.

Lemma 2.2 *If Ω in $\mathbb{R}^n$ is a closed convex set, then x^* is a stationary point for problem* (1.1) *if and only if $x^* \in E[-\nabla f(x^*)]$.*

The proof of this result is a direct consequence of the observation that $x \in E(d)$ if and only if $d \in N(x)$.

Lemma 2.2 shows that any stationary point x^* belongs to the exposed face $E[-\nabla f(x^*)]$. The stationary point x^* may belong to other faces of Ω, but in Section 4 we show that optimization algorithms tend to generate sequences $\{x_k\}$ with $x_k \in E[-\nabla f(x^*)]$ for all k sufficiently large. In this section we develop part of the necessary machinery for this result.

We extend the concept of normal and tangent cones at a point $x \in \Omega$ by defining a normal cone $N(F)$ and a tangent cone $T(F)$ for each $F \in \mathcal{F}$ by

$$N(F) \equiv N(x), \quad T(F) \equiv T(x), \qquad x \in \mathrm{ri}\,(F)\,.$$

This definition relies on the result of Burke and Moré [4] that normal and tangent cones are independent of $x \in \mathrm{ri}\,(F)$ for any $F \in \mathcal{F}$.

The intersection of an arbitrary collection of faces is easily shown to be a face, but the union of faces is not necessarily a face. On the other hand, the following result shows that the union of all faces $F \in \mathcal{F}$ such that $d \in N(F)$ is the face exposed by d.

Lemma 2.3 *If Ω in $\mathbb{R}^n$ is a closed convex set, then*

$$E(d) = \bigcup_{F \in \mathcal{F}(d)} F$$

for any $d \in \mathbb{R}^n$, where $\mathcal{F}(d)$ is the collection of faces $F \in \mathcal{F}$ such that $d \in N(F)$.

Proof. Assume first that $x \in E(d)$. Then $d \in N(x)$, and since Theorem 2.1 guarantees that $x \in \mathrm{ri}\,(F)$ for some $F \in \mathcal{F}$, we have $d \in N(x) = N(F)$. This shows that

$$E(d) \subset \bigcup_{F \in \mathcal{F}(d)} F.$$

We complete the proof by showing that if $F \in \mathcal{F}(d)$ then $F \subset E(d)$. Assume that $x \in F$ for some $F \in \mathcal{F}(d)$. We can choose a sequence $\{x_k\}$ with $x_k \in \mathrm{ri}\,(F)$ such that $\{x_k\}$ converges to x. Since $x_k \in \mathrm{ri}\,(F)$ and $d \in N(F)$, we have $d \in N(x_k)$, and hence, $x_k \in E(d)$. Since the exposed face $E(d)$ is closed and $\{x_k\}$ converges to x, we obtain that $x \in E(d)$ as desired. ∎

The proof of Lemma 2.3 yields the stronger result that $\{\mathrm{ri}\,(F) : F \in \mathcal{F}(d)\}$ is a partition of $E(d)$. This result, however, is not needed in this paper.

Lemma 2.3 also shows that if $d \in N(F)$ for some $F \in \mathcal{F}$, then $F \subset E(d)$. For a general convex set Ω, it is difficult to characterize the vectors $d \in N(F)$ that expose a face F. On the other hand, the following result of Burke [3, Theorem 4.1] provides a complete characterization for polyhedral Ω.

Theorem 2.4 *Let Ω be a polyhedral set in $\mathbb{R}^n$.*

If $x \in \Omega$, then $x \in \operatorname{ri}\{E(d)\}$ if and only if $d \in \operatorname{ri}\{N(x)\}$.

If $F \in \mathcal{F}$, then $F = E(d)$ if and only if $d \in \operatorname{ri}\{N(F)\}$.

Burke [3] proved only the first claim in Theorem 2.4, but it is not difficult to show that both claims are equivalent.

Theorem 2.4 fails if Ω is a general convex set. For example, if Ω is the set in Figure 2.1, and $F = \{(1,0)\}$, then F is not exposed by $d = (1,0)$, but $d \in \operatorname{ri}\{N(F)\}$. Moreover, if $F = E(d)$ for $d = (0,1)$, then d exposes F, but $d \notin \operatorname{ri}\{N(F)\}$. This example shows that the assumption of polyhedrality is required in Theorem A.1 of Burke [3].

The above example also shows that a face may not be exposed by a vector $d \in \operatorname{ri}\{N(F)\}$. In the remainder of this section we show that F is exposed by a vector $d \in \operatorname{ri}\{N(F)\}$ if F is a quasi-polyhedral face. This result is of interest because it leads to a connection with the identification results of Burke and Moré [4] for general convex sets. However, note that the remainder of this section is not needed for the main results in this paper.

Burke and Moré [4] defined quasi-polyhedral faces in terms of the lineality of the tangent cone $T(x)$. Recall that for a cone K in $\mathbb{R}^n$, the lineality $\operatorname{lin}\{K\}$ of the cone K is the largest subspace contained in K. Hence, $\operatorname{lin}\{K\} = K \cap (-K)$.

Definition. A face F of a convex set Ω is quasi-polyhedral if

$$\operatorname{aff}\{F\} = x + \operatorname{lin}\{T(x)\}, \qquad x \in \operatorname{ri}(F).$$

The convex set Ω in Figure (2.1) can be used to illustrate the difference between exposed faces and quasi-polyhedral faces. This set has an infinite number of faces, but the only face that is not exposed by a vector $d \in \operatorname{ri}\{N(F)\}$ is $F = \{(0,1)\}$. Also note that Ω has six quasi-polyhedral faces and that these faces are exposed by any vector $d \in \operatorname{ri}\{N(F)\}$.

Burke and Moré [4] show that the lineality of the tangent cone is the orthogonal complement of the normal cone, that is,

$$\operatorname{lin}\{T(x)\} = N(x)^{\perp},$$

where the orthogonal complement $S^{\perp}$ of a set S is the subspace of vectors v such that $\langle v, w\rangle = 0$ for all $w \in S$. This implies that

$$\operatorname{lin}\{T(x)\}^{\perp} = \operatorname{aff}\{N(x)\}.$$

These results lead to the decomposition

$$T(x) = \operatorname{lin}\{T(x)\} \oplus \operatorname{lin}\{T(x)\}^{\perp} \cap T(x) \tag{2.2}$$

of the tangent cone. This result is valid if $T(x)$ is replaced by a general convex cone K. See, for example, Stoer and Witzgall [23, Theorem 2.10.5].

Theorem 2.5 *A quasi-polyhedral face F of a convex set Ω is exposed by any $d \in \mathrm{ri}\,\{N(F)\}$.*

Proof. The bulk of the proof consists of proving that if $x \in \Omega$, then

$$E(d) = \Omega \cap [x + \mathrm{lin}\,\{T(x)\}], \qquad d \in \mathrm{ri}\,\{N(x)\}\,. \tag{2.3}$$

Given this result, we complete the proof by noting that $F = \Omega \cap \mathrm{aff}\{F\}$ for any face of Ω, and thus

$$F = \Omega \cap [x + \mathrm{lin}\,\{T(x)\}], \qquad x \in \mathrm{ri}\,\{F\}\,,$$

for a quasi-polyhedral face. Hence (2.3) yields that $F = E(d)$ if $d \in \mathrm{ri}\,\{N(x)\} = \mathrm{ri}\,\{N(F)\}$.

We now prove that (2.3) holds; we first show that $\Omega \cap [x + \mathrm{lin}\,\{T(x)\}]$ is a subset of $E(d)$. Choose any y in $\Omega \cap [x + \mathrm{lin}\,\{T(x)\}]$. Then $y - x \in \mathrm{lin}\,\{T(x)\} = N(x)^{\perp}$, and since $d \in N(x)$, we obtain that $\langle d, y - x\rangle = 0$. Since $y \in \Omega$, this implies that $y \in E(d)$.

We now show that $E(d)$ is a subset of $\Omega \cap [x + \mathrm{lin}\,\{T(x)\}]$. Choose any y in $E(d)$. Then $y \in \Omega$, and thus $y - x \in T(x)$. The decomposition (2.2) of the tangent cone $T(x)$ shows that we can write

$$y - x = v_1 + v_2, \qquad v_1 \in \mathrm{lin}\,\{T(x)\}\,, \quad v_2 \in \mathrm{lin}\,\{T(x)\}^{\perp} \cap T(x).$$

If we show that $v_2 = 0$, then $y - x = v_1 \in \mathrm{lin}\,\{T(x)\}$ as desired. We first show that $\langle d, v_2\rangle = 0$. Note that $\langle d, y - x\rangle = 0$ because both y and x belong to $E(d)$, and that $\langle d, v_1\rangle = 0$ because $d \in N(x)$ and $v_1 \in \mathrm{lin}\,\{T(x)\} = N(x)^{\perp}$. Hence, $\langle d, v_2\rangle = 0$.

We claim that $v_2 = 0$ since $d \in \mathrm{ri}\,\{N(x)\}$, $v_2 \in \mathrm{lin}\,\{T(x)\}^{\perp} = \mathrm{aff}\,\{N(x)\}$, and $\langle d, v_2\rangle = 0$. The proof of this claim is not difficult. Since $d \in \mathrm{ri}\,\{N(x)\}$ and $v_2 \in \mathrm{aff}\,\{N(x)\}$, we obtain that $d + \epsilon v_2 \in N(x)$ for $\epsilon > 0$ sufficiently small. We now use that $v_2 \in T(x)$ to obtain that $\langle d + \epsilon v_2, v_2\rangle \le 0$. Since $\langle d, v_2\rangle = 0$, this implies that $v_2 = 0$, as desired. ■

3 Face Geometry

The main result of this section is a partition of $\mathbb{R}^n$ in terms of the face structure of a convex set. We then use this result to extend Lemma 3.3 of Burke and Moré [4]. As we shall see, this extension is crucial to the results on exposing constraints.

Our development requires a few basic properties of projection operators. The projection $P : \mathbb{R}^n \to \Omega$ into a closed convex set Ω is defined by

$$P(x) = \mathrm{argmin}\,\{\|y - x\| : y \in \Omega\}\,,$$

where $\|\cdot\|$ is an inner-product norm. This definition implies that the projection operator P can be characterized in terms of the inner product $\langle\cdot,\cdot\rangle$ by requiring that

$$\langle x - P(x), y - P(x)\rangle \le 0, \qquad y \in \Omega.$$

In terms of normal cones, this characterization requires that

$$x - P(x) \in N[P(x)] \tag{3.1}$$

In particular, note that this characterization implies that

$$P(x+z) = x, \qquad x \in \Omega, \quad z \in N(x). \tag{3.2}$$

We extend this result by proving that

$$P(x+z) = x, \qquad x \in F, \quad z \in N(F) \tag{3.3}$$

for any face $F \in \mathcal{F}$. We prove (3.3) by noting that if $\{x_k\}$ is a sequence in $\operatorname{ri}(F)$ that converges to x, then $z \in N(F) = N(x_k)$, and thus (3.2) implies that $P(x_k + z) = x_k$. The result now follows from the continuity of the projection operator.

We now present two decompositions of $\mathbb{R}^n$ in terms of the face structure of a convex set. The first decomposition extends a result of Goffin [15] for polyhedral Ω.

Theorem 3.1 *If Ω is a closed convex set in $\mathbb{R}^n$, then the collection*

$$\{\operatorname{ri}(F) + N\{F\} : F \in \mathcal{F}\}$$

forms a partition of $\mathbb{R}^n$.

Proof. Let $x \in \mathbb{R}^n$, and note that $x = P(x) + [x - P(x)]$. Certainly $P(x) \in \operatorname{ri}(F)$ for some $F \in \mathcal{F}$, and thus the characterization (3.1) of the projection operator implies that

$$x - P(x) \in N[P(x)] = N(F).$$

Hence, $x = P(x) + [x - P(x)]$ belongs to $\operatorname{ri}(F) + N(F)$, and thus

$$\mathbb{R}^n = \bigcup_{F \in \mathcal{F}} [\operatorname{ri}(F) + N(F)].$$

We now show that this decomposition of $\mathbb{R}^n$ is a partition. Assume that $x_i \in \operatorname{ri}(F_i)$ and $z_i \in N(F_i)$ satisfy $x_1 + z_1 = x_2 + z_2$ for some F_1 and F_2 in $\mathcal{F}$. Then (3.2) implies that

$$x_1 = P(x_1 + z_1) = P(x_2 + z_2) = x_2.$$

Thus, $\operatorname{ri}(F_1) \cap \operatorname{ri}(F_2) \neq \emptyset$. Since the relative interiors of the faces of Ω form a partition of Ω, we must have $F_1 = F_2$ as desired. ■

Theorem 3.1 suggests that it may be possible to decompose $\mathbb{R}^n$ in terms of sets of the form $F + \operatorname{ri}\{N(F)\}$. We now show that this is possible provided Ω is polyhedral. The proof is similar to that of Theorem 3.1, but depends heavily on Theorem 2.4.

Theorem 3.2 *If Ω is a polyhedral set in $\mathbb{R}^n$, then the collection*

$$\{F + \operatorname{ri}\{N(F)\} : F \in \mathcal{F}\}$$

forms a partition of $\mathbb{R}^n$.

Proof. If we define $F = E[x - P(x)]$, then Theorem 2.4 shows that $x - P(x) \in \operatorname{ri}\{N(F)\}$. Moreover, since $x - P(x) \in N[P(x)]$, we obtain that $P(x) \in E[x - P(x)] = F$. Hence,

$$x = P(x) + [x - P(x)] \in F + \operatorname{ri}\{N(F)\}.$$

This proves that

$$\mathbb{R}^n = \bigcup_{F \in \mathcal{F}} [F + \operatorname{ri}\{N(F)\}].$$

We now prove that this decomposition of $\mathbb{R}^n$ is a partition. Assume that $x_i \in F_i$ and $z_i \in \operatorname{ri}\{N(F_i)\}$ satisfy $x_1 + z_1 = x_2 + z_2$ for some F_1 and F_2 in $\mathcal{F}$. Then (3.2) implies that

$$x_1 = P(x_1 + z_1) = P(x_2 + z_2) = x_2.$$

Thus, $z_1 = z_2$. Since $z_i \in \operatorname{ri}\{N(F_i)\}$, Theorem 2.4 shows that $F_i = E(z_i)$, and thus $z_1 = z_2$ implies that $F_1 = F_2$ as desired. ■

Theorem 3.2 does not hold for general convex sets Ω. For example, if Ω is the convex set shown in Figure 2.1, then any point of the form $(0, \xi)$ with $\xi > 1$ does not belong to a set of the form $F + \operatorname{ri}\{N(F)\}$ for $F \in \mathcal{F}$. On the other hand, Theorem 3.1 shows that $(0, \xi)$ with $\xi > 1$ must belong to a set of the form $\operatorname{ri}(F) + N\{F\}$; a computation shows that $F = \{(0, 1)\}$.

We proved Theorems 3.1 and 3.2 because of the interest in partitions of $\mathbb{R}^n$ in terms of the faces of a convex set. In the remainder of this section, however, we only need to know that if Ω is polyhedral then the collection $\{F + N(F) : F \in \mathcal{F}\}$ covers $\mathbb{R}^n$. Clearly, both Theorems 3.1 and 3.2 yield this result.

In the proof of the lemma below we make use of the result that if Ω is polyhedral then Ω has a finite number of faces. This result can be found, for example, in Rockafellar [22, Theorem 19.1]. We also need to know that if Ω is closed and convex, then any face of Ω is closed. For this result, see Stoer and Witzgall [23, Theorem 3.6.6].

Lemma 3.3 *Assume that Ω is a polyhedral set in $\mathbb{R}^n$. If $x \in \Omega$ and $d \in N(x)$, then*

$$x + d \in \operatorname{int}\{K(d)\},$$

where

$$K(d) = \bigcup_{F \in \mathcal{F}(d)} [F + N(F)],$$

and $\mathcal{F}(d)$ is defined in Lemma 2.3.

Proof. The first step in the proof is to show that if $F \in \mathcal{F}$, then $F + N(F)$ is closed. The definition of a normal cone shows that $N(F)$ is closed, and since Ω is closed, each face F is closed. Now let $\{y_k\}$ be a sequence in $F + N(F)$ that converges, and let $y_k = x_k + z_k$, where $x_k \in F$ and $z_k \in N(F)$. Since (3.3) implies that

$$x_k = P(x_k + z_k) = P(y_k),$$

and $\{y_k\}$ converges, it follows that $\{x_k\}$ converges. Moreover, since $y_k = x_k + z_k$, the sequence $\{z_k\}$ also converges. We now use that $x_k \in F$ and $z_k \in N(F)$, to conclude that $\{x_k\}$ converges to some $x \in F$ and $\{z_k\}$ converges to some $z \in N(F)$. Hence, $\{y_k\}$ converges to $x + z$ in $F + N(F)$. This shows that $F + N(F)$ is closed.

The next step in the proof is technical; we need to show that $x + d$ does not belong to the set

$$L(d) = \bigcup_{F \notin \mathcal{F}(d)} [F + N(F)].$$

Theorem 3.1 shows that $x + d \in F + N(F)$ for some $F \in \mathcal{F}$. Let $y \in F$ and $z \in N(F)$ be such that $x + d = y + z$. Then (3.3) implies that

$$x = P(x + d) = P(y + z) = y,$$

and thus $d = z \in N(F)$. Hence, $F \in \mathcal{F}(d)$. This shows that $x + d \notin L(d)$ as desired.

We have shown that $F + N(F)$ is closed for each $F \in \mathcal{F}$. Since Ω is polyhedral, there is a finite number of faces, and thus $L(d)$ is closed. Moreover, since $x + d \notin L(d)$, there is an open set U such that $x + d \in U$ and $U \cap L(d) = \emptyset$. Now note that Theorem 3.1 implies that $K(d) \cup L(d) = \mathbb{R}^n$, and thus $U \subset K(d)$. Hence,

$$x + d \in U \subset \operatorname{int}\{K(d)\},$$

as desired. ■

Lemmas 2.1 and 3.3 are the ingredients needed to prove the main result of this section.

Theorem 3.4 *Let Ω be a polyhedral set in $\mathbb{R}^n$. Assume that $\{x_k\}$ and $\{d_k\}$ are sequences in $\mathbb{R}^n$ with $x_k \in \Omega$ and $d_k \in N(x_k)$. If $\{x_k\}$ converges to x^* and $\{d_k\}$ converges to d^*, then $x_k \in E(d^*)$ for all k sufficiently large.*

Proof. Lemma 3.3 implies that $x^* + d^* \in \operatorname{int}\{K(d^*)\}$ so that $x_k + d_k \in K(d^*)$ for all k sufficiently large. Since $d_k \in N(x_k)$,

$$x_k = P(x_k + d_k) \in P[K(d^*)]$$

by virtue of (3.2). Now note that (3.3) and Lemma 2.1 imply that

$$P[K(d^*)] \subset \left\{ \bigcup_{F \in \mathcal{F}(d^*)} P[F + N(F)] \right\} = \left\{ \bigcup_{F \in \mathcal{F}(d^*)} F \right\} = E(d^*),$$

and thus $x_k \in E(d^*)$ for all k sufficiently large. ■

Burke and Moré [4] proved a similar result for general Ω by assuming that $x^* \in \operatorname{ri}\{F\}$ for a quasi-polyhedral face F and $d^* \in \operatorname{ri}\{N(F)\}$. The assumption that $d^* \in \operatorname{ri}\{N(F)\}$ is a nondegeneracy assumption that is avoided in Theorem 3.4 by assuming that Ω is polyhedral.

Theorem 3.4 fails if Ω is a general closed convex set even if we assume that $x^* \in \operatorname{ri}\{F\}$ for a quasi-polyhedral face F. For example, consider the set Ω shown in Figure 2.1. The sequence $\{x_k\}$ defined by

$$x_k = \left(\sin(\frac{\pi}{k}), \cos(\frac{\pi}{k})\right)$$

belongs to Ω for $k > 1$, and $d_k = x_k \in N(x_k)$. However, $\{d_k\}$ converges to $d^* = (0,1)$, and $E(d^*) = \{(0,1)\}$. Thus, $x_k \notin E(d^*)$ for any $k > 1$.

4 Exposing Constraints

Previous results on the identification properties of algorithms in the neighborhood of a stationary point either assumed linear independence of the active constraint normals or nondegeneracy of the stationary point. In this section we avoid these restrictions.

Definition. Given a convex set Ω, a stationary point x^* is nondegenerate if

$$-\nabla f(x^*) \in \operatorname{ri}\{N(x^*)\}. \tag{4.1}$$

This nondegeneracy condition is due to Dunn [10]. An advantage of this definition is that it does not make any linear independence assumptions on the constraints. Also note that Burke and Moré [4] proved that if Ω is polyhedral, then x^* is nondegenerate if and only if there is a set of positive Lagrange multipliers. This result implies that this definition of nondegeneracy can be viewed as a generalization of the standard strict complementarity condition.

The main result of this section is an extension of the following identification result of Burke and Moré [4]. For this result recall that $F(x)$ is the unique face of Ω that contains x in the relative interior.

Theorem 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on the closed, convex set Ω. If $\{x_k\}$ is a sequence in Ω that converges to a nondegenerate stationary point x^* of* (1.1), *then*

$$\lim_{k\to\infty} P_{T(x_k)}[-\nabla f(x_k)] = 0 \tag{4.2}$$

if and only if there is a $k_0 > 0$ such that

$$x_k \in \operatorname{ri}\{F(x^*)\}, \qquad k \geq k_0.$$

An advantage of Theorem 4.1 is that it is independent of the representation of Ω. For applications, however, it is necessary to express Theorem 4.1 in terms of a specific representation of Ω. Note, in particular, that if Ω is the polyhedral set defined by the set of linear constraints

$$\Omega = \{x \in \mathbb{R}^n : \langle c_j, x\rangle \geq \delta_j,\ 1 \leq j \leq m\}, \tag{4.3}$$

for some vectors $c_j \in \mathbb{R}^n$ and scalars δ_j, then Calamai and Moré [6] proved that the projected gradient that appears in (4.2) can be expressed in the familiar form

$$P_{T(x)}[-\nabla f(x)] = -\nabla f(x) + \sum_{j \in \mathcal{A}(x)} \lambda_j c_j, \tag{4.4}$$

where the set of active constraints is defined by

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

and λ_j for $j \in \mathcal{A}(x)$ solves the bound constrained linear least squares problem

$$\min\left\{ \left\| \nabla f(x) - \sum_{j \in \mathcal{A}(x)} \lambda_j c_j \right\| : \lambda_j \geq 0 \right\}.$$

Also note that representation (4.4) of the projected gradient is unique even if the active constraints c_j with $j \in \mathcal{A}(x)$ are linearly dependent.

Theorem 4.1 characterizes the limiting behavior of many algorithms with respect to the face structure of Ω. For example, Burke and Moré [4] prove that the sequential quadratic programming algorithm and the gradient projection algorithm generate sequences that satisfy (4.2). Consequently, if the stationary point x^* is nondegenerate, the sequences generated by these algorithms eventually enter and remain in the relative interior of the face $F(x^*)$. This can be extremely useful knowledge in the design of algorithms for the solution of (1.1).

Note that Theorem 4.1 gives information on the identification of the active constraints. This is based on the observation of Burke and Moré [4] that

$$\text{ri}\,\{F(x^*)\} = \{x \in \Omega : \langle c_j, x\rangle = \delta_j,\ j \in \mathcal{A}(x^*),\ \ \langle c_j, x\rangle > \delta_j,\ j \notin \mathcal{A}(x^*)\}.$$

Thus, Theorem 4.1 shows that if $\{x_k\}$ is a sequence in Ω that converges to a nondegenerate stationary point x^* of (1.1), then the sequence $\{P_{T(x_k)}[-\nabla f(x_k)]\}$ converges to zero if and only if there is a $k_0 > 0$ such that

$$\mathcal{A}(x_k) = \mathcal{A}(x^*), \qquad k \geq k_0.$$

The following example of Stephen Wright shows that this result fails in degenerate cases.

Example. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(\xi_1, \xi_2) = \tfrac{1}{2}\left(\xi_1^2 + \mu \xi_2^2\right)$$

for some $\mu > 1$, and define $\Omega \subset \mathbb{R}^2$ by

$$\Omega = \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_1 \geq \mu^2 \xi_2,\ \xi_1 \geq -\xi_2 \right\}.$$

Note that f is a strictly convex quadratic and that $x^* = 0$ is the global minimizer in Ω. A computation shows that the steepest descent iterates are generated by

$$\begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}_+ = \frac{\mu - 1}{\xi_1^2 + \mu^3 \xi_2^2} \begin{pmatrix} \mu^2 \xi_1 \xi_2^2 \\ -\xi_1^2 \xi_2 \end{pmatrix}.$$

This implies, in particular, that all the iterates are feasible and that

$$\frac{\xi_1^{(k+2)}}{\xi_2^{(k+2)}} = \frac{\xi_1^{(k)}}{\xi_2^{(k)}}.$$

Hence, every other iterate of the steepest descent method lies in the same ray from the origin. Thus, if the initial iterate satisfies either the constraint $\xi_1 = \mu^2 \xi_2$ or the constraint $\xi_1 = -\xi_2$, the iterates of the steepest descent method oscillate between the two constraints that define Ω. Since these iterates are feasible, the steepest descent iterates coincide with the gradient projection iterates. This implies that in this example,

$$\mathcal{A}(x_{k+1}) \neq \mathcal{A}(x_k), \quad \mathcal{A}(x_k) \neq \mathcal{A}(x^*), \qquad k \geq k_0.$$

Of course, in this case x^* is degenerate because $\nabla f(x^*) = 0$.

We now consider the case when x^* is degenerate and show that the nondegeneracy assumption in Theorem 4.1 can be dropped if we replace $\operatorname{ri}\{F(x^*)\}$ by $E[-\nabla f(x^*)]$.

Theorem 4.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on the polyhedral set Ω, and assume that $\{x_k\}$ is a sequence in Ω that converges to a stationary point x^* of* (1.1)*. Then*

$$\lim_{k \to \infty} P_{T(x_k)} [-\nabla f(x_k)] = 0$$

if and only if there is a $k_0 > 0$ such that

$$x_k \in E[-\nabla f(x^*)], \qquad k \geq k_0.$$

Proof. The proof depends on the Moreau decomposition (see, for example, Lemma 2.2 of Zarantonello [26]) and Theorem 3.4. The Moreau decomposition of $-\nabla f$ shows that

$$-\nabla f(x_k) = -P_{T(x_k)} [-\nabla f(x_k)] + P_{N(x_k)}[-\nabla f(x_k)].$$

In particular,

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

where $d_k \in N(x_k)$. Hence, if the sequence $\{P_{T(x_k)}[-\nabla f(x_k)]\}$ converges to zero, then $\{d_k\}$ converges to $-\nabla f(x^*)$, and thus Theorem 3.4 shows that $x_k \in E[-\nabla f(x^*)]$ for all k sufficiently large.

Conversely, assume that $x_k \in E[-\nabla f(x^*)]$ for all k sufficiently large. Since $x \in E(d)$ if and only if $d \in N(x)$, we have that there is a $k_0 > 0$ such that $-\nabla f(x^*) \in N(x_k)$ for all $k \geq k_0$. Since the Moreau decomposition implies that

$$\left\| P_{T(x_k)}[-\nabla f(x_k)] \right\| = \min \{ \|\nabla f(x_k) + d\| : d \in N(x_k) \},$$

we obtain that

$$\left\| P_{T(x_k)}[-\nabla f(x_k)] \right\| \leq \|\nabla f(x_k) - \nabla f(x^*)\|.$$

This yields the desired result. ∎

In the remainder of this section we examine various consequences of Theorem 4.2. We first use Burke's result (Theorem 2.4) to show that Theorem 4.2 is an extension of Theorem 4.1 for polyhedral Ω.

Theorem 4.3 *If Ω is a polyhedral set, then x^* is a nondegenerate stationary point for problem* (1.1) *if and only if $x^* \in \operatorname{ri}\{E[-\nabla f(x^*)]\}$.*

Proof. This result is a direct consequence of Theorem 2.4. ∎

We can show that Theorem 4.2 is an extension of Theorem 4.1 by noting that if x^* is a nondegenerate stationary point for problem (1.1), then Theorem 4.3 shows that x^* is in $\operatorname{ri}\{E[-\nabla f(x^*)]\}$, and thus any $x \in E[-\nabla f(x^*)]$ sufficiently close to x^* also belongs to $\operatorname{ri}\{E[-\nabla f(x^*)]\}$.

We now show that if the polyhedral set Ω has the specific representation (4.3), then the face $E[-\nabla f(x^*)]$ can be expressed in terms of the positive Lagrange multipliers at x^*.

Theorem 4.4 *If Ω is the polyhedral set in $\mathbb{R}^n$ defined by* (4.3) *and*

$$\nabla f(x^*) = \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* c_j, \qquad \lambda_j^* \geq 0, \tag{4.5}$$

for some stationary point x^* of problem (1.1)*, then*

$$E[-\nabla f(x^*)] = \left\{ x \in \Omega : \langle c_j, x \rangle = \delta_j,\ \lambda_j^* > 0 \right\}.$$

Proof. Note that

$$\langle \nabla f(x^*), x \rangle = \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* \langle c_j, x \rangle \geq \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* \delta_j$$

for any $x \in \Omega$. Thus, $\langle \nabla f(x^*), x \rangle$ is minimized only when $\langle c_j, x \rangle = \delta_j$ for $\lambda_j^* > 0$. ∎

If the active constraint normals are linearly dependent then there is an infinite set of Lagrange multipliers that satisfy (4.5). Nevertheless, Theorem 4.4 shows that the set

$$\left\{ x \in \Omega : \langle c_j, x \rangle = \delta_j, \ \lambda_j^* > 0 \right\}$$

is independent of the choice of Lagrange multipliers.

Theorem 4.5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on the polyhedral set Ω defined by* (4.3). *Assume that x^* is a stationary point of* (1.1) *and that the Lagrange multipliers λ_i^* satisfy* (4.5). *If $\{x_k\}$ is a sequence in Ω that converges to x^*, then*

$$\lim_{k \to \infty} P_{T(x_k)} [-\nabla f(x_k)] = 0 \tag{4.6}$$

if and only if there is a $k_0 > 0$ such that

$$\{ i \in \mathcal{A}(x^*) : \lambda_i^* > 0 \} \subset \mathcal{A}(x_k), \qquad k \geq k_0. \tag{4.7}$$

Proof. Theorem 4.4 shows that

$$x \in E[-\nabla f(x^*)] \quad \Longleftrightarrow \quad \{ i \in \mathcal{A}(x^*) : \lambda_i^* > 0 \} \subset \mathcal{A}(x), \quad x \in \Omega.$$

Thus, the result is a direct consequence of Theorem 4.2. ■

This result has immediate applications to several algorithms. For example, Calamai and Moré [6] show that the gradient projection method generates sequences that satisfy (4.6). Hence, Theorem 4.5 implies that if the sequence $\{x_k\}$ generated by the gradient projection method converges to x^*, then (4.7) holds.

Closely related convergence results for the gradient projection method are due to Bertsekas [1, 2], Gafni and Bertsekas [12, 13], and Dunn [9]. However, in these papers it is not shown that (4.6) holds; instead, it is shown that any limit point of the sequence $\{x_k\}$ is stationary.

Theorem 4.5 also has application to the class of trust region methods for bound constrained optimization problems proposed by Conn, Gould, and Toint [7, 8]. For this algorithm Lescrenier [17] proved that (4.7) holds. Hence, Theorem 4.5 implies that (4.6) holds. This result is of interest because the measure

$$\nu(x; f) = \left\| P_{T(x)} [-\nabla f(x)] \right\| \tag{4.8}$$

is a natural criterion for terminating the algorithm. This is clear from expression (4.4) for the projected gradient. Moreover, note that $\nu(x; \cdot)$ is scale invariant in the sense that $\nu(x; \alpha f) = \alpha \nu(x, f)$ for any $\alpha > 0$. Other criteria, for example,

$$\nu(x; f) = \| P[x - \nabla f(x)] - x \|$$

do not have this important property.

Finally, Theorem 4.5 has implications to the class of trust region methods for general linearly constrained methods analyzed by Burke, Moré, and Toraldo [5]. These results will be reported in a later paper.

5 Strictly Binding Constraints

We have stressed the geometric viewpoint of Theorem 4.2 because this viewpoint leads to results that are independent of the representation of Ω. On the other hand, the viewpoint of Theorem 4.5 is needed because it is closely related to computational issues. In this section we show that it is possible to define a set of strictly binding constraints $\mathcal{B}_s^*$ at x^*, independent of the choice of Lagrange multipliers. This leads, in particular, to a version of Theorem 4.5 in terms of $\mathcal{B}_s^*$.

We will show that the set of strictly binding constraints $\mathcal{B}_s^*$ can be defined in terms of the active set of the exposed face $E[-\nabla f(x^*)]$.

Definition. If Ω is a polyhedral set and F is a face of Ω, then

$$\mathcal{A}(F) = \mathcal{A}(x), \qquad x \in \mathrm{ri}\,\{F\}$$

is the active set of F.

We justify this definition by showing that the active set $\mathcal{A}(x)$ is independent of $x \in \mathrm{ri}\,\{F\}$. Assume that Ω is defined by (4.3), and choose any $x \in \mathrm{ri}\,\{F\}$. If $y \in F$, then

$$x_\lambda = x + \lambda(y - x) \in F$$

for some $\lambda < 0$. Hence, if $i \in \mathcal{A}(x)$, then

$$\delta_i \le \langle c_i, x_\lambda \rangle = (1-\lambda)\langle c_i, x\rangle + \lambda\langle c_i, y\rangle = (1-\lambda)\delta_i + \lambda\langle c_i, y\rangle.$$

Since $\lambda < 0$, this implies that $\langle c_i, y\rangle \le \delta_i$. Thus, $i \in \mathcal{A}(y)$. We have shown that $\mathcal{A}(x) \subset \mathcal{A}(y)$ if $x \in \mathrm{ri}\,\{F\}$ and $y \in F$. Hence, $\mathcal{A}(x) = \mathcal{A}(y)$ if x and y belong to $\mathrm{ri}\,\{F\}$.

This definition of $\mathcal{A}(F)$ is direct and intuitive. We now establish an alternative characterization of $\mathcal{A}(F)$. For the proof of this result we need to know that

$$-d \in \mathrm{ri}\,\{N(x)\} \quad \Longleftrightarrow \quad d = \sum_{i \in \mathcal{A}(x)} \lambda_i c_i, \quad \lambda_i > 0. \tag{5.1}$$

This characterization of $\mathrm{ri}\,\{N(x)\}$ is due to Burke and Moré [4, Lemma 3.2].

Lemma 5.1 *If Ω is a polyhedral set, then*

$$F = \{x \in \Omega : \langle c_i, x\rangle = \delta_i,\ i \in \mathcal{A}(F)\}$$

for any face F of Ω.

Proof. Every face of a polyhedron is exposed (see, for example, Theorem 2.4.12 of Stoer and Witzgall [23]), and thus $F = E(-d)$ for some $d \in \mathbb{R}^n$. Hence, Theorem 2.4 implies that $-d \in \operatorname{ri}\{N(F)\}$, and thus $d \in \operatorname{ri}\{N(x)\}$ for any $x \in \operatorname{ri}\{F\}$. We now appeal to the characterization (5.1), and conclude that there are $\lambda_i > 0$ such that

$$d = \sum_{i \in \mathcal{A}(x)} \lambda_i c_i, \qquad \lambda_i > 0.$$

A computation now shows that

$$E[-d] = \{x \in \Omega : \langle c_i, x\rangle = \delta_i,\ i \in \mathcal{A}(x)\}.$$

Since $\mathcal{A}(x) = \mathcal{A}(F)$ and $F = E[-d]$, this establishes the result. ∎

This result characterizes $\mathcal{A}(F)$ as the set of active constraints that are essential for F. In view of this result, it is natural to investigate the properties of the active set of the exposed face $E[-\nabla f(x^*)]$.

Definition. If Ω is a polyhedral set and x^* is a stationary point of (1.1), then

$$\mathcal{B}_s^* = \mathcal{A}\Big(E[-\nabla f(x^*)]\Big)$$

is the set of strictly binding constraints at x^*.

This definition requires justification because strictly binding constraints are usually associated with positive Lagrange multipliers; the following result provides this justification.

Theorem 5.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on the polyhedral set Ω defined by* (4.3)*. If x^* is a stationary point of* (1.1) *then*

$$\nabla f(x^*) = \sum_{j \in \mathcal{B}_s^*} \lambda_j^* c_j, \qquad \lambda_j^* > 0, \tag{5.2}$$

for some set of positive Lagrange multipliers λ_i^*.

Proof. Theorem 2.4 implies that $-\nabla f(x^*) \in \operatorname{ri}\{N(F)\}$ for $F = E[-\nabla f(x^*)]$. Hence, $-\nabla f(x^*) \in \operatorname{ri}\{N(x)\}$ for any $x \in \operatorname{ri}\{F\}$. Since $\mathcal{A}(x) = \mathcal{B}_s^*$ for $x \in \operatorname{ri}\{F\}$, the characterization (5.1) yields the result. ∎

Theorem 5.2 justifies the definition of $\mathcal{B}_s^*$ as the set of strictly binding constraints. This result is of interest because it does not make any nondegeneracy or linear independence assumptions. Note that if x^* is nondegenerate then Theorem 4.3 shows that $x^* \in \operatorname{ri}\{E[-\nabla f(x^*)]\}$, and thus $\mathcal{B}_s^* = \mathcal{A}(x^*)$. Hence, for nondegenerate x^*, Theorem 5.2 is a direct consequence of (5.1) and the definition of nondegeneracy.

Theorem 5.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on the polyhedral set Ω defined by* (4.3)*. If x^* is a stationary point of* (1.1) *and $\{\lambda_i^*\}$ is any set of Lagrange multipliers, then the following three statements are equivalent.*

1. $x \in E[-\nabla f(x^*)]$.

2. $\{i \in \mathcal{A}(x^*) : \lambda_i^* > 0\} \subset \mathcal{A}(x)$ *for* $x \in \Omega$.

3. $\mathcal{B}_s^* \subset \mathcal{A}(x)$ *for* $x \in \Omega$.

Proof. Theorem 4.4 shows that the first two statements are equivalent. The equivalence of the first and third statements is obtained by noting that Lemma 5.1 implies that

$$E[-\nabla f(x^*)] = \{x \in \Omega : \langle c_i, x \rangle = \delta_i, \ i \in \mathcal{B}_s^*\}.$$

■

This result shows that Theorem 4.5 can be phrased in terms of $\mathcal{B}_s^*$. Another interesting consequence of Theorem 5.3 is obtained by choosing any $x \in \text{ri}\{E[-\nabla f(x^*)]\}$ in the second statement in Theorem 5.3. This shows that

$$\{i \in \mathcal{A}(x^*) : \lambda_i^* > 0\} \subset \mathcal{B}_s^*$$

for any set of Lagrange multipliers $\{\lambda_i^*\}$. Hence, the indices in $\mathcal{B}_s^*$ identify a maximal set of positive Lagrange multipliers.

6 Second-Order Sufficiency Conditions

We have shown that optimization algorithms tend to generate iterates such that

$$x_k \in E[-\nabla f(x^*)]$$

for all k sufficiently large. Another important component in the analysis of an optimization algorithm is to show that the iterates converge under suitable conditions. This type of result usually requires the assumption that the iterates $\{x_k\}$ have an isolated limit point x^*. This assumption is satisfied, for example, if x^* is an *isolated* stationary point, that is, there is a neighborhood $S(x^*)$ such that x^* is the only stationary point in $S(x^*) \cap \Omega$. In this section we show that if the Hessian is positive definite in the cone generated by $E[-\nabla f(x^*)] - x^*$, then the stationary point x^* is isolated.

The assumption that x^* is an isolated stationary point can be guaranteed by imposing second-order conditions on f. The following result uses a version of the second-order sufficiency conditions that is appropriate for problem (1.1) with a general convex set Ω.

Theorem 6.1 *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be continuously differentiable on a closed convex set* Ω *and twice differentiable at a point* x^* *in* Ω. *If* x^* *is a stationary point of problem* (1.1) *and*

$$\langle \nabla f(x^*), w \rangle = 0, \quad w \in T(x^*), \quad w \neq 0 \quad \Longrightarrow \quad \langle w, \nabla^2 f(x^*) w \rangle > 0, \tag{6.1}$$

then x^* *is an isolated stationary point of* f.

Theorem 6.1 is a special case of a result of Robinson [21, Theorem 2.4]. The proof of Theorem 6.1 is not difficult in our setting. Note that if x^* is not an isolated stationary point, then there is a sequence $\{x_k\}$ of stationary points converging to x^*. In particular,

$$\langle \nabla f(x^*), x_k - x^* \rangle \geq 0, \qquad \langle \nabla f(x_k), x^* - x_k \rangle \geq 0. \tag{6.2}$$

This implies that

$$\langle \nabla f(x_k) - \nabla f(x^*), x_k - x^* \rangle \leq 0. \tag{6.3}$$

Hence, if w is a limit point of the sequence $\{w_k\}$ defined by

$$w_k = \frac{x_k - x^*}{\|x_k - x^*\|},$$

then (6.2) shows that $\langle \nabla f(x^*), w \rangle = 0$, while (6.3) shows that $\langle w, \nabla^2 f(x^*) w \rangle \leq 0$. Moreover, since $w_k \in T(x^*)$ and $\|w_k\| = 1$, we also have that $w \in T(x^*)$ and that $\|w\| = 1$. This contradicts (6.1).

For a polyhedral Ω, condition (6.1) coincides with the standard second-order sufficiency conditions. This observation follows by noting that if Ω is the set defined by (4.3) and

$$\nabla f(x^*) = \sum_{j \in \mathcal{A}(x^*)} \lambda_j^* c_j, \qquad \lambda_j^* \geq 0,$$

then

$$\begin{aligned} &\{w \in \mathbb{R}^n : \langle \nabla f(x^*), w \rangle = 0, \quad w \in T(x^*)\} = \\ &\qquad \left\{ w \in \mathbb{R}^n : \langle c_j, w \rangle \geq 0, \ j \in \mathcal{A}(x^*), \quad \langle c_j, w \rangle = 0, \ \lambda_j^* > 0 \right\}. \end{aligned}$$

For a general convex Ω, condition (6.1) has an advantage over the standard second-order sufficiency conditions because it is independent of the representation of Ω. On the other hand, an example of Burke, Moré, and Toraldo [5] shows that condition (6.1) does not take into account the curvature of Ω, and thus differs from the standard second-order sufficiency conditions.

If x^* is a nondegenerate stationary point, then (6.1) can be expressed in terms of $N(x^*)^\perp$ where for any set S the orthogonal complement $S^\perp$ of S is the subspace of vectors v such that $\langle v, w \rangle = 0$ for all $w \in S$. Indeed, Burke, Moré,, and Toraldo [5] prove that if x^* is a nondegenerate stationary point, then

$$N(x^*)^\perp = \{ \ w \in T(x^*) : \langle \nabla f(x^*), w \rangle = 0\}. \tag{6.4}$$

These results are familiar when Ω is the polyhedral set defined by (4.3) because

$$N(x^*)^\perp = \{v \in \mathbb{R}^n : \langle c_j, v \rangle = 0, \ j \in \mathcal{A}(x^*)\}.$$

We now show that the second-order sufficiency condition can be expressed in terms of the exposed face $E[-\nabla f(x^*)]$. In this result, cone$\{S\}$ is the cone spanned by the set S, that is, the set of vectors αw for some $\alpha \geq 0$ and $w \in S$.

Theorem 6.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on a polyhedral Ω and twice differentiable at a point x^* in Ω. If x^* is a stationary point of problem* (1.1), *then*

$$\{w \in T(x^*) : \langle \nabla f(x^*), w \rangle = 0\} = \text{cone}\,\{E[-\nabla f(x^*)] - x^*\}\,. \tag{6.5}$$

Moreover, if

$$w \in \text{cone}\,\{E[-\nabla f(x^*)] - x^*\}\,, \quad w \neq 0 \quad \Longrightarrow \quad \langle w, \nabla^2 f(x^*) w \rangle > 0, \tag{6.6}$$

then x^* is an isolated stationary point of f.

Proof. The result follows from Theorem 6.1 if we establish (6.5). We first show that

$$\text{cone}\,\{E[-\nabla f(x^*)] - x^*\} \subset \{w \in T(x^*) : \langle \nabla f(x^*), w \rangle = 0\}\,.$$

If $w \in \text{cone}\,\{E[-\nabla f(x^*)] - x^*\}$, then $w = \alpha(x - x^*)$ for some $x \in E[-\nabla f(x^*)]$. Thus, since $\langle \nabla f(x^*), \cdot \rangle$ is constant on $E[-\nabla f(x^*)]$, we obtain that $\langle \nabla f(x^*), w \rangle = 0$. Moreover, since $x \in \Omega$, it is clear that $w \in T(x^*)$. For the reverse inclusion assume that $\langle \nabla f(x^*), w \rangle = 0$ for some $w \in T(x^*)$. Since $w \in T(x^*)$ and Ω is polyhedral, $x^* + \alpha w \in \Omega$ for all $\alpha > 0$ sufficiently small. Moreover, since $\langle \nabla f(x^*), w \rangle = 0$, it follows that $x^* + \alpha w \in E[-\nabla f(x^*)]$. Thus, $w \in \text{cone}\,\{E[-\nabla f(x^*)] - x^*\}$ as desired. ■

We have already noted that if x^* is nondegenerate, then (6.4) holds. This implies, in particular, that

$$\{w \in T(x^*) : \langle \nabla f(x^*), w \rangle = 0\}$$

is a subspace. In view of (6.5), we have shown that if x^* is nondegenerate then the second-order condition (6.1) reduces to the assumption that $\nabla^2 f(x^*)$ is positive definite on the subspace $\text{aff}\,\{E[-\nabla f(x^*)] - x^*\}$. In the context of algorithms, the fact that (6.5) is a subspace is extremely helpful. However, the same effect can be obtained without assuming nondegeneracy if one is willing to strengthen the second-order sufficiency condition.

Definition. Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable at $x^* \in \Omega$. The strong second-order sufficiency condition for problem (1.1) is satisfied at a stationary point x^* if $\nabla^2 f(x^*)$ is positive definite on the subspace $\text{aff}\,\{E[-\nabla f(x^*)] - x^*\}$, that is,

$$w \in \text{aff}\,\{E[-\nabla f(x^*)] - x^*\}\,, \quad w \neq 0 \quad \Longrightarrow \quad \langle w, \nabla^2 f(x^*) w \rangle > 0. \tag{6.7}$$

If the strong second-order sufficiency condition (6.7) holds, then the standard second-order sufficiency condition (6.1) must also hold because

$$\text{cone}\,\{E[-\nabla f(x^*)] - x^*\} \subset \text{aff}\,\{E[-\nabla f(x^*)] - x^*\}\,.$$

The following example shows that the converse can fail even if f is quadratic.

Example. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

$$f(\xi_1, \xi_2) = \xi_1^2 + 4\xi_1\xi_2 + \xi_2^2,$$

and let $\Omega \subset \mathbb{R}^2$ be given by

$$\Omega = \left\{ (\xi_1, \xi_2) \in \mathbb{R}^2 : \xi_1 \geq 0,\ \xi_2 \geq 0 \right\}.$$

For this problem $x^* = (0, 0)$ is the global minimum of f over Ω. Note that $\nabla f(x^*) = 0$, and thus $E[-\nabla f(x^*)] = \Omega$. Hence,

$$\operatorname{cone}\{E[-\nabla f(x^*)] - x^*\} = \mathbb{R}^2_+, \qquad \operatorname{aff}\{E[-\nabla f(x^*)] - x^*\} = \mathbb{R}^2.$$

A computation shows that the standard second-order sufficiency condition (6.1) holds at x^*. However, it is clear that the strong second-order sufficiency condition (6.7) is not satisfied at x^* because $\nabla^2 f(x^*)$ is not positive definite.

Other authors have used the term strong second-order sufficiency condition to mean that

$$\langle c_j, w \rangle = 0,\ \lambda_j^* > 0, \quad w \neq 0 \quad \Longrightarrow \quad \langle w, \nabla^2 f(x^*) w \rangle > 0. \tag{6.8}$$

For example, Gay [14], Lescrenier [17], and Robinson [20] use this condition in their analysis of the convergence behavior of algorithms. A disadvantage of condition (6.8) is that it depends on the representation of Ω and the choice of multipliers. The following example shows that condition (6.8) is stronger than (6.7) even if f is a quadratic.

Example. Let $f : \mathbb{R}^3 \to \mathbb{R}$ be given by

$$f(\xi_1, \xi_2, \xi_3) = \xi_3^2 - \xi_2^2 + \xi_3,$$

and let $\Omega = \{x \in \mathbb{R}^3 : \langle c_i, x \rangle \geq 0,\ 1 \leq i \leq 4\}$ be the cone with vertex at the origin, where

$$c_1 = e_1, \quad c_2 = e_2, \quad c_3 = e_3 - e_1, \quad c_4 = e_3 - e_2.$$

For this problem $x^* = (0, 0, 0)$ is the global minimum of f over Ω. A computation shows that $E[-\nabla f(x^*)] = \{x^*\}$, and thus the strong second-order condition (6.7) holds. However, condition (6.8) does not hold if we choose $\lambda_2^* = \lambda_4^* = 0$.

This example shows that condition (6.8) is stronger than (6.7) for some set of multipliers. On the other hand, the following result shows that there is a set of multipliers such that condition (6.8) is equivalent to (6.7).

Theorem 6.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable on the polyhedral set Ω defined by* (4.3)*. If x^* is a stationary point of* (1.1)*, then*

$$\operatorname{aff}\{E[-\nabla f(x^*)] - x^*\} = \{v \in \mathbb{R}^n : \langle c_j, v \rangle = 0,\ j \in \mathcal{B}_s^*\},$$

where $\mathcal{B}_s^*$ is the set of strictly binding constraints defined in Section 5.

Proof. If $v \in E[-\nabla f(x^*)] - x^*$, then $\langle \nabla f(x^*), v \rangle = 0$, because $\langle \nabla f(x^*), \cdot \rangle$ is constant on $E[-\nabla f(x^*)]$. Moreover, since v is a feasible direction, $\langle c_j, v \rangle \geq 0$ for $j \in \mathcal{A}(x^*)$. Now recall that Theorem 5.2 guarantees that (5.2) holds. Hence,

$$0 = \langle \nabla f(x^*), v \rangle = \sum_{j \in \mathcal{B}_s^*} \lambda_j^* \langle c_j, v \rangle \geq 0.$$

This implies that $\langle c_j, v \rangle = 0$ when $j \in \mathcal{B}_s^*$. Thus, we have shown that

$$E[-\nabla f(x^*)] - x^* \subset \{ v \in \mathbb{R}^n : \langle c_j, v \rangle = 0, \ j \in \mathcal{B}_s^* \}.$$

Hence,

$$\text{aff}\{E[-\nabla f(x^*)] - x^*\} \subset \{ v \in \mathbb{R}^n : \langle c_j, v \rangle = 0, \ j \in \mathcal{B}_s^* \}. \tag{6.9}$$

For the reverse inclusion choose any $x_0 \in \text{ri}\{E[-\nabla f(x^*)]\}$. If $\langle c_j, v \rangle = 0$ for $j \in \mathcal{B}_s^*$, then v is a feasible direction because $\mathcal{B}_s^* = \mathcal{A}(x_0)$. Hence $x_0 + \alpha v \in \Omega$ for all $\alpha > 0$ sufficiently small. Moreover, (5.2) implies that

$$\langle \nabla f(x^*), x + \alpha v \rangle = \langle \nabla f(x^*), x \rangle.$$

This shows that $x_0 + \alpha v \in E[-\nabla f(x^*)]$. We have established that

$$\{ v \in \mathbb{R}^n : \langle c_j, v \rangle = 0, \ j \in \mathcal{B}_s^* \} \subset E[-\nabla f(x^*)] - x_0.$$

Hence,

$$dim \{ v \in \mathbb{R}^n : \langle c_j, v \rangle = 0, \ j \in \mathcal{B}_s^* \} \leq dim \{E[-\nabla f(x^*)]\}.$$

Since we have already established (6.9), this completes the proof. ■

7 Algorithms

Theorem 4.5 is an important tool for the convergence analysis of optimization algorithms. In this section we examine the use of Theorem 4.5 in the GPCG algorithm of Moré and Toraldo [19]. In particular, we show that the convergence condition of the GPCG algorithm is satisfied in a finite number of iterations even when the solution is degenerate.

The GPCG algorithm uses a combination of the gradient projection algorithm and the conjugate gradient algorithm to solve large-scale problems of the form

$$\min\{q(x) : l \leq x \leq u\}. \tag{7.1}$$

The GPCG algorithm uses the conjugate gradient method to explore the active set

$$\mathcal{A}(x) = \{i : x_i \in \{l_i, u_i\}\} \tag{7.2}$$

defined by the current iterate. Once this exploration is completed, the gradient projection method is used to choose a new active set. The convergence properties of algorithm GPCG are summarized in the following result of Moré and Toraldo [19].

Theorem 7.1 *Let $q : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex quadratic. If $\{x_k\}$ is the sequence generated by algorithm GPCG for problem (7.1), then either $\{x_k\}$ terminates at the solution x^* in a finite number of iterations, or $\{x_k\}$ converges to the solution x^* of problem (7.1). If the solution x^* of problem (7.1) satisfies the nondegeneracy condition*

$$\partial_i q(x^*) \neq 0, \qquad i \in \mathcal{A}(x^*),$$

then algorithm GPCG terminates at the solution x^* in a finite number of iterations.

This convergence result is not entirely satisfactory when the solution x^* is degenerate, because it does not show that the convergence test is satisfied in a finite number of iterations. Given a starting point x_0 in the feasible region $\Omega = \{x \in \mathbb{R}^n : l \leq x \leq u\}$ and a tolerance $\tau > 0$, the convergence test of the GPCG algorithm requires that

$$\|\nabla_\Omega q(x_k)\| \leq \tau \|\nabla q(x_0)\|, \tag{7.3}$$

where $\nabla_\Omega q$ is defined by

$$[\nabla_\Omega q(x)]_i = \begin{cases} \partial_i q(x) & \text{if } x_i \in (l_i, u_i) \\ \min\{\partial_i q(x), 0\} & \text{if } x_i = l_i \\ \max\{\partial_i q(x), 0\} & \text{if } x_i = u_i \end{cases}.$$

This convergence test is closely related to (4.6) because

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

In particular, if (4.6) holds, then the convergence test (7.3) is satisfied after a finite number of iterations. In the remainder of this section we show that (4.6) holds.

Given the current iterate x_k, algorithm GPCG explores the active set defined by the current iterate by computing an approximate minimizer of the subproblem

$$\min\{q(x_k + d) : d_i = 0,\ i \in \mathcal{A}(x_k)\}. \tag{7.4}$$

Given an approximate minimizer d_k of subproblem (7.4), algorithm GPCG uses a projected search to choose a search parameter α_k such that $q(x_{k+1}) < q(x_k)$, where

$$x_{k+1} = P(x_k + \alpha_k d_k)$$

and P is the projection into the feasible region Ω. Details on the projected search can be found in the paper of Moré and Toraldo [19]. For this paper it is necessary to note only that since the i-th coordinate is zero when $i \in \mathcal{A}(x_k)$, we obtain that

$$\mathcal{A}(x_k) \subset \mathcal{A}(x_{k+1}) \tag{7.5}$$

holds whenever algorithm GPCG explores the active set defined by the current iterate.

The approximate minimizer d_k of subproblem (7.4) is obtained by first noting that if $i_1, \ldots, i_{m_k}$ are the indices of the *free* variables, that is, those variables with indices outside of $\mathcal{A}(x_k)$, then subproblem (7.4) is equivalent to the unconstrained subproblem

$$\min\{q_k(w) : w \in \mathbb{R}^{m_k}\}, \tag{7.6}$$

where $q_k : \mathbb{R}^{m_k} \to \mathbb{R}$ is defined by $q_k(w) \equiv q(x_k + Z_k w)$, and Z_k is the matrix in $\mathbb{R}^{n \times m_k}$ whose j-th column is the i_j-th column of the identity matrix in $\mathbb{R}^{n \times n}$. Given the starting point $w_0 = 0$ in $\mathbb{R}^{m_k}$, the conjugate gradient algorithm generates a sequence of iterates $\{w_{j_k}\}$. The approximate solution of subproblem (7.4) is then $d_k = Z_k w_{j_k}$ for some $j_k \geq 0$.

If the iterate x_{k+1} generated by the conjugate gradient method appears to have identified the active set defined by the solution, then algorithm GPCG explores this active set further. The decision to continue the conjugate gradient method is based on the observation that if $\mathcal{A}(x) = \mathcal{A}(x^*)$, then the *binding set*

$$\mathcal{B}(x) = \{i : x_i = l_i \text{ and } \partial_i q(x) \geq 0, \text{ or } x_i = u_i \text{ and } \partial_i q(x) \leq 0\}$$

agrees with the active set $\mathcal{A}(x)$. Thus, if the conjugate gradient method produces an iterate x_{k+1} such that $\mathcal{B}(x_{k+1}) = \mathcal{A}(x_{k+1})$, then algorithm GPCG continues to use the conjugate gradient method to explore this active set. The finite termination properties of the conjugate gradient algorithm show that at most m_k iterations are needed before the conjugate gradient algorithm finds a solution of subproblem (7.4); at this point there would be no need to explore this active set further.

Once the conjugate gradient algorithm has explored an active set, algorithm GPCG uses the gradient projection method

$$y_{k+1} = P\left[y_k - \alpha_k \nabla q(y_k)\right]$$

with $y_0 = x_k$ to select a new active set. The GPCG algorithm sets $x_{k+1} = y_{j_k}$ for some $j_k \geq 0$. Additional details on the implementation of the gradient projection method can be found in the paper of Moré and Toraldo [19].

We now use Theorem 4.5 to prove that (4.6) holds for the iterates generated by algorithm GPCG. The proof consists of showing that condition (4.7) holds. For the bound constrained problem (7.1), condition (4.7) is just that

$$\mathcal{B}_s^* \equiv \{i \in \mathcal{A}(x^*) : \partial_i q(x^*) \neq 0\} \subset \mathcal{A}(x_k), \qquad k \geq k_0,$$

where $\mathcal{B}_s^*$ is the set of strictly binding constraints at x^* defined in Section 5.

Theorem 7.2 *Let $q : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex quadratic. If $\{x_k\}$ is the sequence generated by algorithm GPCG for problem* (7.1)*, then either $\{x_k\}$ terminates at the solution x^* in a finite number of iterations, or there is an index k_0 such that the convergence test* (7.3) *is satisfied for all $k \geq k_0$.*

Proof. Theorem 7.1 shows that either $\{x_k\}$ terminates at the solution x^* in a finite number of iterations, or $\{x_k\}$ converges to x^*. The result follows if $\{x_k\}$ terminates at the solution x^*, so we assume that $\{x_k\}$ converges to x^*. Let $\mathcal{K}_{GP}$ be the set of indices k such that x_{k+1} is generated by the gradient projection method. Theorem 5.2 of Calamai and Moré [6] shows that

$$\lim_{k \in \mathcal{K}_{GP}, k \to \infty} P_{T(x_{k+1})}[-\nabla q(x_{k+1})] = 0.$$

Hence, Theorem 4.6 shows that

$$\mathcal{B}_s^* \subset \mathcal{A}(x_{k+1})$$

if $k \in \mathcal{K}_{GP}$ if k is sufficiently large. Moreover, since (7.5) holds whenever the conjugate gradient method is used to explore the current active set, we immediately obtain that

$$\mathcal{B}_s^* \subset \mathcal{A}(x_k)$$

holds for all sufficiently large indices k. Hence, (4.7) holds, and thus Theorem 4.5 shows that (4.6) holds as desired. ■

Acknowledgments. We thank Stephen Wright for his interesting example in Section 4 of a degenerate optimization problem where the active set does not settle down. We are also grateful to Gail Pieper for her gentle comments on the manuscript.

References

[1] D. P. Bertsekas, *On the Goldstein–Levitin–Polyak gradient projection method*, IEEE Trans. Automat. Control, 21 (1976), pp. 174–184.

[2] ——, *Projected Newton methods for optimization problems with simple constraints*, SIAM J. Control Optim., 20 (1982), pp. 221–246.

[3] J. V. Burke, *On the identification of active constraints II: The nonconvex case*, SIAM J. Numer. Anal., 27 (1990), pp. 1081–1102.

[4] J. V. Burke and J. J. Moré, *On the identification of active constraints*, SIAM J. Numer. Anal., 25 (1988), pp. 1197–1211.

[5] J. V. Burke, J. J. Moré, and G. Toraldo, *Convergence properties of trust region methods for linear and convex constraints*, Math. Programming, 47 (1990), pp. 305–336.

[6] P. H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, Math. Programming, 39 (1987), pp. 93–116.

[7] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Global convergence of a class of trust region algorithms for optimization problems with simple bounds*, SIAM J. Numer. Anal., 25 (1988), pp. 433–460.

[8] ——, *Testing a class of methods for solving minimization problems with simple bounds on the variables*, Math. Comp., 50 (1988), pp. 399–430.

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

[10] ——, *On the convergence of projected gradient processes to singular critical points*, J. Optim. Theory Appl., 55 (1987), pp. 203–216.

[11] ——, *A projected Newton method for minimization problems with nonlinear inequality constraints*, Numer. Math., 53 (1988), pp. 377–410.

[12] E. M. Gafni and D. P. Bertsekas, *Convergence of a gradient projection method*, Report LIDS-P-1201, Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Cambridge, Massachusetts, 1982.

[13] ——, *Two-metric projection methods for constrained optimization*, SIAM J. Control Optim., 22 (1984), pp. 936–964.

[14] D. M. Gay, *A trust region approach to linearly constrained optimization*, in Numerical Analysis, D. F. Griffiths, ed., Lecture Notes in Mathematics 1066, Springer-Verlag, 1984, pp. 72–105.

[15] J. L. Goffin, *The relaxation method for solving systems of linear inequalities*, Math. Oper. Res., 5 (1980), pp. 388–414.

[16] C. T. Kelley and E. W. Sachs, *Mesh independence of the gradient projection method for optimal control problems*, SIAM J. Control Optim., 30 (1992), pp. 477–493.

[17] M. Lescrenier, *Convergence of trust region algorithms for optimization with bounds when strict complementarity does not hold*, SIAM J. Numer. Anal., 28 (1991), pp. 476–495.

[18] J. J. Moré, *Trust regions and projected gradients*, in Systems Modelling and Optimization, M. Iri and K. Yajima, eds., Lecture Notes in Control and Information Sciences 113, Springer-Verlag, 1988, pp. 1–13.

[19] J. J. Moré and G. Toraldo, *Algorithms for bound constrained quadratic programming problems*, Numer. Math., 55 (1989), pp. 377–400.

[20] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.

[21] ——, *Generalized equations and their solutions, part II: Applications to nonlinear programming*, Math. Programming Stud., 19 (1982), pp. 200–221.

[22] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[23] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag, 1970.

[24] S. J. Wright, *Implementing proximal point methods for linear programming*, J. Optim. Theory Appl., 65 (1990), pp. 531–554.

[25] ——, *Identifiable surfaces in constrained optimization*, SIAM J. Control Optim., (1992). to appear.

[26] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory*, in Contributions to Nonlinear Functional Analysis, E. H. Zarantonello, ed., Academic Press, 1971, pp. 237–424.