SIAM J. NUMER. ANAL.
Vol. 27, No. 4, pp. 1081–1102, August 1990

016

ON THE IDENTIFICATION OF ACTIVE CONSTRAINTS II: THE NONCONVEX CASE*

JIM BURKE†

Abstract. In this paper the results of Burke and Moré [5] on the identification of active constraints are extended to the nonconvex constrained nonlinear programming problem. The approach is motivated by the geometric structure of a certain polyhedral convex "linearization" of the constraint region at each iteration. As in Burke and Moré [5] questions of constraint identification are couched in terms of the faces of these polyhedra. The main result employs a nondegeneracy condition due to Dunn [7] and the linear independence condition to obtain a characterization of those algorithms that identify the optimal active constraints in a finite number of iterations. The role of the linear independence condition is carefully examined and it is argued that it is required within the context of a first-order theory of constraint identification. In conclusion, the characterization theorem is applied to the Wilson–Han–Powell sequential quadratic programming algorithm [*J. Optim. Theory Appl.*, 22 (1977), pp. 297–309], [*Proceedings of the* 1977 *Dundee Conference on Numerical Analysis*, Springer-Verlag, Berlin, 1977], and [*A simplicial algorithm for concave programming*, Ph.D. thesis, Graduate School of Business Administration, Harvard University, Boston, 1963] and Fletcher's *QL* algorithm [*Practical Methods for Optimization*, Vol. 2, John Wiley, New York, 1981], [*Math. Programming Study*, 17 (1982), pp. 67–76].

Key words. active constraints, linear independence condition, Mangasarian–Fromowitz constraint qualification, nondegeneracy, strict complementarity, constrained optimization

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

1. Introduction. In this paper we examine the problem of active constraint identification for the constrained nonlinear programming problem

$$\begin{array}{lll} \text{NLP:} & \text{minimize } f(x) & \\ & \text{subject to} & g_i(x) \le 0 \quad i = 1, 2, \cdots, s \\ & & g_i(x) = 0 \quad i = s+1, \cdots, m \end{array}$$

where it is always assumed that the functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$ for $i = 1, \cdots, m$ belong to $C^1[\mathbb{R}^n, \mathbb{R}]$ the set of functions mapping $\mathbb{R}^n$ into $\mathbb{R}$ having continuous Fréchet derivatives. Specifically, given a sequence $\{x^k\} \subset \mathbb{R}^n$ converging to a local solution $\overline{x} \in \Omega$ of NLP, where

$$\Omega := \{x : g_i(x) \le 0, \quad i = 1, \cdots, s, \; g_i(x) = 0, \quad i = s+1, \cdots, m\}$$

is the feasible region for NLP, we want to establish whether or not the iterates x^k identify the set of indices

$$I(\bar{x}) := \{i \in \{1, \cdots, s\} : g_i(\bar{x}) = 0\}$$

of the active constraints at $\overline{x}$ for all k sufficiently large. The ability to characterize those sequences that possess this identification property has significant applications in the design and analysis of algorithms for solving NLP. For example, we could

*Received by the editors July 18, 1988; accepted for publication (in revised form) September 20, 1989. This work was supported in part by the National Science Foundation under grant DMS-8803206.

†Department of Mathematics, GN–50, University of Washington, Seattle, Washington 98195.

apply these results toward the development of hybrid algorithms wherein a procedure that does not possess this identification property can be augmented by occasionally computing an iterate of a method that does possess the identification property. In [5], Burke and Moré study this problem for the case in which Ω is polyhedral convex (or, more generally, when Ω is convex and the solution lies on a face of Ω (§4) which is "quasi-polyhedral" [5, Def. 2.5]) and the sequence $\{x^k\}$ lies entirely in Ω. In this case they show that if $\overline{x} \in \Omega$ is a nondegenerate local solution to NLP in the sense of Dunn [7] (see §3) with Ω polyhedral, then for any sequence $\{x^k\} \subset \Omega$ converging to $\overline{x}$ we have $I(\overline{x}) = I(x^k)$ for all k sufficiently large if and only if the projection of $-\nabla f(x^k)$ onto the tangent cone to Ω at x^k (see §§3 and 6) converges to the origin. In the present paper it is not assumed that Ω is convex, nor is it assumed that the sequence $\{x^k\}$ lies entirely in Ω, consequently a somewhat different approach is required. Nevertheless, the results we obtain have the same flavor as those of [5].

The key to the approach taken in this paper is the notion of a linearization of the set Ω. The linearization for Ω that we consider is well studied in the literature [1], [14], [15], [17]. This linearization, denoted $L\Omega(x,r)$, is defined in §2 as a multifunction from $\mathbb{R}^n \times \mathbb{R}^m$ into $\mathbb{R}^n$. For each $(x,r) \in \mathbb{R}^n \times \mathbb{R}^m$ $L\Omega(x,r)$ is a, possibly empty, polyhedral convex subset of $\mathbb{R}^n$. Moreover, for each $x \in \mathbb{R}^n$ there is always a choice of $r \in \mathbb{R}^m$ such that $L\Omega(x,r) \neq \emptyset$. In particular, if $x \in \Omega$, then $x \in L\Omega(x,0)$. Using the linearization $L\Omega(\cdot,\cdot)$ we develop a theory of constraint identification paralleling that given in [5]. Of particular note is Theorem 6.2. In this result we assume that $\overline{x} \in \Omega$ is a stationary point for NLP at which the strict complementary slackness and linear independence conditions (see Definition 6.1) are satisfied. It is then shown that if $\{(x^k, r^k, y^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ is a sequence converging to $(\overline{x}, 0, \overline{x})$ with $y^k \in L\Omega(x^k, r^k)$ for all $k = 1, 2, \cdots$, then the constraints defining $L\Omega(x^k, r^k)$ that are active at y_k will identify the constraints $I(\overline{x})$ for all k sufficiently large if and only if the projection of $-\nabla f(x^k)$ onto the tangent cone to $L\Omega(x^k, r^k)$ at y^k converges to the origin. Although this result is not surprising it does provide a simple unification of the known theory on the identification of active constraints.

The outline of the paper is as follows. In §2 we introduce the linearization $L\Omega(\cdot,\cdot)$. In §3 optimality and nondegeneracy conditions for NLP are given in terms of $L\Omega(\cdot,\cdot)$. A theory paralleling that of [5] is then developed. The first step is to recall certain facts concerning the geometry of convex polyhedra. This is done in §4. In §5 these geometric concepts are related to various notions of identification for the faces of a convex polyhedron and in §6 we prove our main results. In our analysis the linear independence (LI) condition (Definition 6.1) plays a key role. This is unfortunate since the LI condition is a rather severe restriction in the presence of inequality constraints. Nonetheless, we show that this condition is in a sense unavoidable in the context of the first-order theory of constraint identification developed herein. However, it is hoped that the present work will provide the basis for a second-order theory wherein the LI condition is not required. The paper is concluded in §7 by applying the results of the previous sections to generalizations of the sequential quadratic programming algorithm of Wilson [19], Han [10], and Powell [13], and the Q.L. (or S_{ℓ_1} QP) algorithm of Fletcher [8],]9]. In this regard our results complement the recent work of Wright [20], [21] on the problem of active constraint identification for these algorithms.

The notation that we use is standard; however, a few words about our use of norms is appropriate. The symbol $\|\cdot\|$ always denotes a norm on $\mathbb{R}^n$ and $\mathbb{B}$ denotes the closed unit ball associated with this norm. The norm dual to $\|\cdot\|$ is denoted $\|\cdot\|_*$ and is given by

$$\|y\|_* := \sup\{\langle x, y\rangle : x \in \mathbb{B}\}$$

where $\langle x, y\rangle := \sum_{i=1}^{n} x_i y_i$. The closed unit ball associated with $\|\cdot\|_*$ is given by $\mathbb{B}^0 = \{y : \langle x, y\rangle \leq 1 \;\; \forall x \in \mathbb{B}\}$. The ℓ_2 or Euclidean norm plays a special role in our discussion. It is defined by

$$\|x\|_2 := [\langle x, x\rangle]^{1/2},$$

and its associated closed unit ball is denoted by $\mathbb{B}_2$. For a given norm $\|\cdot\|$ and a given set $S \subset \mathbb{R}^n$ we define the distance function for S by

$$\operatorname{dist}(x|S) := \inf\{\|x - s\| : s \in S\}.$$

If the norm employed is the ℓ_2 norm we specify this by writing $\operatorname{dist}_2(x|S)$.

The polar set of a set $S \subset \mathbb{R}^n$ is defined as

$$S^0 := \{x^* : \langle x^*, x\rangle \leq 1 \text{ for all } x \in S\}.$$

If S is a cone, we can show that

$$S^0 := \{x^* : \langle x^*, x\rangle \leq 0 \text{ for all } x \in S\}.$$

Given a convex subset C of $\mathbb{R}^n$ and a point $x \in C$ the tangent cone to C at x is defined as

$$T(x|C) := c\ell\{\lambda(c - x) : \lambda \geq 0,\;\; c \in C\},$$

where $c\ell(S)$ denotes the closure of the set S, and the normal cone to C at x is

$$N(x|C) := \{x^* : \langle x^*, c - x\rangle \leq 0 \text{ for all } c \in C\}.$$

We also have the identity

$$T(x|C)^0 = N(x|C).$$

For a function $f : \mathbb{R}^n \to \mathbb{R}$ and a set $S \subset \mathbb{R}^n$, we write

$$\arg\min\{f(x) : x \in S\} := \{\overline{x} \in S : f(\overline{x}) = \min\{f(x) : x \in S\}\},$$

where this set may be empty if no such $\overline{x}$ exists. Finally we make use of the following convention regarding finite sums. If $I = \emptyset$, then we define

$$\sum_I a_i = 0$$

for any choice of argument a_i where 0 is the origin in the vector space of the appropriate dimension.

2. The linearization of Ω. Since most algorithms for solving NLP do not maintain feasibility, our approach to questions of constraint identification must be able to obtain information about constraint activity from nonfeasible points. In order to do this we make use of the notion of a linearization of the constraint region Ω and work with the linearizations of Ω rather than Ω itself. The linearization that we consider is well studied in the literature [1], [14], [15], [17] and is given by the multifunction $L\Omega : \mathbb{R}^{n+m} \Longrightarrow \mathbb{R}^n$ where

$$L\Omega(x, r) := \left\{ y \;\middle|\; \begin{array}{l} g_i(x) + \nabla g_i(x)^T(y - x) \leq r_i,\;\; i = 1, \cdots, s \\ g_i(x) + \nabla g_i(x)^T(y - x) = r_i,\;\; i = s+1, \cdots, m \end{array} \right\}$$

for each $(x, r) \in \mathbb{R}^n \times \mathbb{R}^m$. Observe that for each $(x, r) \in \mathbb{R}^n \times \mathbb{R}^m$ the set $L\Omega(x, r)$ is a, possibly empty, convex polyhedron. Moreover, for each $x \in \mathbb{R}^n$, $r \in \mathbb{R}^m$ can be chosen so that $L\Omega(x, r)$ is nonempty.

For each $y \in L\Omega(x, r)$ we call the set

$$A(x, r, y) := \{i \in \{1, 2, \cdots, s\} : g_i(x) + \nabla g_i(x)^T(y - x) = r_i\}$$

the set of active constraints of $L\Omega(x, r)$ at y. The indices $\{s, \cdots, m\}$ are not included in $A(x, r, y)$ since they are always known to be active. For $\overline{x} \in \Omega$, the set of active constraints of Ω at $\overline{x}$ is just $A(\overline{x}, 0, \overline{x})$. The following lemma describes the local relationship between $A(x, r, y)$ and $A(\overline{x}, 0, \overline{x})$ at points $\overline{x} \in \Omega$.

LEMMA 2.1. *If $\overline{x} \in \Omega$, there exist neighborhoods U of $(\overline{x}, 0)$ in $\mathbb{R}^n \times \mathbb{R}^m$ and V of $\overline{x}$ in $\mathbb{R}^n$ such that for all (x, r) in U and y in V we have*

$$A(x, r, y) \subset A(\overline{x}, 0, \overline{x}).$$

Proof. Since $g = (g_1, \cdots, g_m)^T$ belongs to $C^1[\mathbb{R}^n, \mathbb{R}^m]$, there are neighborhoods U of $(\overline{x}, 0)$ and V of $\overline{x}$ such that

$$g_i(x) + \nabla g_i(x)^T(y - x) < r_i$$

for $i \notin A(\overline{x}, 0, \overline{x})$ whenever $(x, r) \in U$ and $y \in V$. The lemma clearly holds for this choice of U and V. □

Lemma 2.1 illustrates that any sequence $\{(x_i, r_i, y_i)\}$ converging to a point $(\overline{x}, 0, \overline{x})$ with $\overline{x} \in \Omega$ identifies subsets of $A(\overline{x}, 0, \overline{x})$ in a finite number of iterations. The question is whether these subsets are meaningful in the context of NLP. In order to address this issue we need to examine the geometry of the sets $L\Omega(x, r)$ vis-à-vis the optimality conditions for NLP. Our first step in this direction is to review the basic results on first-order optimality conditions for NLP.

3. Review of optimality conditions for NLP and nondegeneracy. Given a nonempty closed set S contained in $\mathbb{R}^n$ and a point $\overline{x}$ in S, we denote by $\pi_S(\overline{x})$ the set of all vectors $v \in \mathbb{R}^n$ such that

$$\|v\|_2 = \text{dist}_2(\overline{x} + v | S);$$

that is

$$\pi_S(\overline{x}) := \{v \in \mathbb{R}^n : \overline{x} \in \arg\min\{\|\overline{x} + v - y\|_2 : y \in S\}\}.$$

The Clarke normal cone to S at $\overline{x}$ is then defined to be the set

$$N(\overline{x}|S) := \left\{\lambda \lim_{i\to\infty} \frac{v^i}{\|v^i\|_2} : \lambda \geq 0, \{x^i\} \subset S, v^i \in \pi_S(x^i), x^i \to \overline{x}, v^i \to 0\right\}$$

(see [6]). This notation is consistent with that employed for the normal cone to a convex set since the two notions of normality coincide in this case [6, Prop. 2.4.4]. In [6], it is shown that the condition

$$-\nabla f(\overline{x}) \in N(\overline{x}|\Omega) \tag{3.1}$$

is a first-order necessary condition for optimality in NLP at points $\overline{x} \in \Omega$. Thus we say that $\overline{x}$ is a stationary point for NLP if (3.1) holds. By contrast a point $\overline{x} \in \Omega$ is said to be a Kuhn–Tucker point for NLP if

$$-\nabla f(\overline{x}) \in N(\overline{x}|L\Omega(\overline{x},0)). \tag{3.2}$$

The cones $N(\overline{x}|\Omega)$ and $N(\overline{x}|L\Omega(\overline{x},0))$ are related by the containment

$$N(\overline{x}|\Omega) \supset N(\overline{x}|L\Omega(\overline{x},0)) \tag{3.3}$$

[6, p. 56]. This containment can be strict as is illustrated in the following example.

Example. Consider the set

$$\Omega := \{x \in \mathbb{R} \mid x^2 \leq 0,\ 0 \leq x\}.$$

Since $\Omega = \{0\}$, we have $N(0|\Omega) = \mathbb{R}$. However, $N(0|L\Omega(0,0)) = \mathbb{R}_-$. Consequently,

$$N(0|L\Omega(0,0)) \subset \operatorname{int}[N(0|\Omega)].$$

Hence, in general, (3.2) is not necessary for optimality. Conditions under which (3.2) is necessary for optimality are called constraint qualifications. The most important of these constraint qualifications is the Mangasarian–Fromowitz constraint qualification here after denoted by MFCQ.

DEFINITION 3.1. We say that the MFCQ is satisfied at a point $\overline{x} \in \Omega$ if the only set of scalars u_i for $i = 1, \cdots, m$ satisfying

$$\begin{aligned} 0 &= \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) \\ 0 &= \sum_{i=1}^{s} u_i g_i(\overline{x}), \\ 0 &\leq u_i \quad \text{for } i = 1, \cdots, s, \end{aligned}$$

is $u_i = 0$ for $i = 1, \cdots, m$.

The above definition is known as the dual form of the Mangasarian–Fromowitz constraint qualification and is equivalent to the form that is usually given in the literature [12]. Clarke [6, Cor. 2, p. 56] establishes the following result.

THEOREM 3.1. *If the* MFCQ *is satisfied at* $\overline{x} \in \Omega$, *then*

$$N(\overline{x}|\Omega) = N(\overline{x}|L\Omega(\overline{x},0)).$$

Hence the MFCQ yields the necessity of condition (3.2). We make strong use of the MFCQ in the following section.

Another concept that we will find useful is that of nondegeneracy. Following Dunn [7], we say that $\overline{x} \in \Omega$ is a nondegenerate stationary point for NLP if

$$-\nabla f(\overline{x}) \in \operatorname{ri}(N(\overline{x}|\Omega)), \tag{3.4}$$

where for any subset $S \subset \mathbb{R}^n$ the relative interior of S, ri S, is the interior of S relative to the affine hull of S,

$$\operatorname{aff}(S) := \left\{ \sum_{i=1}^{\ell} \alpha_i x_i \;\middle|\; \begin{array}{l} \ell \in \{1,2,\cdots,\}, \{\alpha_i\}_{i=1}^{\ell} \subset \mathbb{R}, \{x_i\}_{i=1}^{\ell} \subset S \\ \text{and } \sum_{i=1}^{\ell} \alpha_i = 1 \end{array} \right\}.$$

If the MFCQ is also satisfied at $\overline{x}$, then (3.4) can be written as

$$-\nabla f(\overline{x}) \in \mathrm{ri}(N(\overline{x}|L\Omega(\overline{x},0))). \tag{3.5}$$

Now since $L\Omega(\overline{x},0)$ is polyhedral we have

(3.6)

$$N(\overline{x}|L\Omega(\overline{x},0)) = \left\{ \sum_{i\in A(\overline{x},0,\overline{x})} u_i\nabla g_i(\overline{x}) + \sum_{i=s+1}^{m} u_i\nabla g_i(\overline{x}) : u_i \geq 0, i \in A(\overline{x},0,\overline{x}) \right\}$$

and so by [5, Lem. 3.2]

(3.7)

$$\mathrm{ri}(N(\overline{x}|L\Omega(\overline{x},0))) = \left\{ \sum_{i\in A(\overline{x},0,\overline{x})} u_i\nabla g_i(\overline{x}) + \sum_{i=s+1}^{m} u_i\nabla g_i(\overline{x}) : u_i > 0, i \in A(\overline{x},0,\overline{x}) \right\}.$$

Consequently the nondegeneracy condition (3.5) is closely related to the so-called strict complementary slackness condition. This relationship is carefully examined in [5].

4. Faces of convex sets. As in [5], properties of faces of convex sets play an important role in our analysis. In this section we briefly provide the necessary background and notation.

Recall that a nonempty convex subset $\widehat{C}$ of a closed convex set C in $\mathbb{R}^n$ is said to be a face of C if every convex subset of C whose relative interior meets $\widehat{C}$ is contained in $\widehat{C}$ (e.g., see [16, §18]). In fact, the relative interiors of the faces of C form a partition of C [16, Thm. 18.2]. Thus every point $x \in C$ can be associated with a unique face of C denoted by

$$F(x|C)$$

such that $x \in \mathrm{ri}\,(F(x|C))$. A face $\widehat{C}$ of C is said to be exposed if there is a vector $x^* \in \mathbb{R}^n$ such that $\widehat{C} = E(x^*|C)$ where

$$E(x^*|C) := \arg\max\{\langle x^*, y\rangle : y \in C\}.$$

The vector x^* is said to expose the face $E(x^*|C)$. It is well known and elementary to show that every face $\widehat{C}$ of a polyhedron is exposed and that the exposing vectors are precisely the elements of $\mathrm{ri}\,(N(x|C))$ for any $x \in \mathrm{ri}\,\widehat{C}$. Here $N(x|C)$ is the normal cone to C at x defined in the previous section. Since C is convex the normal cone has representation

$$N(x|C) = \{x^* \in \mathbb{R}^n : \langle x^*, y - x\rangle \leq 0 \;\; \forall y \in C\}.$$

THEOREM 4.1. *Let $C \subset \mathbb{R}^n$ be a nonempty polyhedron and let $x \in C$. Then*

$$E(x^*|C) = F(x|C)$$

if and only if $x^* \in \mathrm{ri}\,(N(x|C))$.

The proof of Theorem 4.1 is contained in Appendix A wherein a complete characterization of exposed faces and their exposing vectors is given for general convex sets.

PROPOSITION 4.1. *Let C be a nonempty closed convex subset of $\mathbb{R}^n$ and let $x \in C$.*

(1) *If $y_1, y_2 \in \mathrm{ri}\,(F(x|C))$, then $N(y_1|C) = N(y_2|C)$.*

(2) *If* $y \in F(x|C)$, *then* $N(x|C) \subset N(y|C)$.

(3) *The following conditions are equivalent:*

(a) $x^* \in N(x|C)$.

(b) $x \in E(x^*|C)$.

(c) $F(x|C) \subset E(x^*|C)$.

(4) *If* C *is polyhedral, then the condition*

$$\mathrm{ri}(N(x|C)) \cap N(z|C) \neq \emptyset$$

for some $z \in C$ *implies that*

$$F(z|C) \subset F(x|C).$$

Remark. Part (1) above indicates that it makes sense to speak of the normal and tangent cones to a face of C. Thus, as in [5, Def. 2.4], if $\widehat{C}$ is a face of C, define

$$N(\widehat{C}|C) := N(x|C) \quad \text{and} \quad T(\widehat{C}|C) := T(x|C)$$

for any $x \in \mathrm{ri}\,(\widehat{C})$. For example,

$$T(C|C) = \mathrm{aff}\,(C) - x$$

for any $x \in C$ where $\mathrm{aff}\,(C)$ is the affine hull of C.

Proof. (1) See [5, Thm. 2.3].

(2) If $y \in F(x|C)$, then for all $\varepsilon > 0$ there is a $z \in \mathrm{ri}\,(F(x|C))$ such that $\|y-z\| < \varepsilon$. Let $x^* \in N(x|C)$ and so by (1) $x^* \in N(z|C)$ also. Hence, for $\overline{y} \in C$

$$\langle x^*, \overline{y} - y\rangle = \langle x^*, \overline{y} - z\rangle + \langle x^*, z - y\rangle \leq \varepsilon \|x^*\|_*.$$

Letting $\varepsilon \to 0$ we obtain the result.

(3) By definition $x^* \in N(x|C)$ if and only if $\langle x^*, z\rangle \leq \langle x^*, x\rangle$ for all $z \in C$. Hence the equivalence of (a) and (b) is obvious. Now since $x \in F(x|C)$, (c) implies (b). On the other hand, if $x^* \in N(x|C)$, then

$$\langle x^*, y - x\rangle \leq 0 \quad \text{for all } y \in C,$$

and so

$$\langle x^*, y - x\rangle = 0 \quad \text{for all } y \in F(x|C),$$

since by [5, inclusion (2.4) and Lem. 2.7],

$$\mathrm{aff}\,[F(x|C)] - x \subset \lim(T(x|C)) = N(x|C)^{\perp},$$

where $N(x|C)^{\perp} := \{y : \langle y, z\rangle = 0 \text{ for all } z \in N(x|C)\}$. Consequently, $F(x|C) \subset E(x^*|C)$.

(4) Let $x^* \in \mathrm{ri}\,(N(x|C)) \cap N(z|C)$. By Theorem 4.1, $F(x|C) = E(x^*|C)$, and, by Part (3), $F(z|C) \subset E(x^*|C)$. □

Since the normal cone structure of polyhedral convex sets plays such a significant role in our study we close this section with a characterization of this structure. The verification of this result can be found in many places including [5].

PROPOSITION 4.2. *Let C be a polyhedral convex subset of $\mathbb{R}^n$ with representation*

$$C := \{x \in \mathbb{R}^n : \langle c_i, x\rangle \le \gamma_i, \quad i = 1, \cdots, s, \langle c_i, x\rangle = \gamma_i, \quad i = s+1, \cdots, m\}. \tag{4.1}$$

Then for $x \in C$ we have

$$N(x|C) := \left\{ \sum_{i \in I(x)} u_i c_i + \sum_{i=s+1}^{n} u_i c_i : u_i \ge 0, \quad i \in I(x) \right\} \tag{4.2}$$

and

$$\mathrm{ri}(N(x|C)) := \left\{ \sum_{i \in I(x)} u_i c_i + \sum_{i=s+1}^{m} u_i c_i : u_i > 0, \quad i \in I(x) \right\} \tag{4.3}$$

where

$$I(x) := \{i \in \{1, \cdots, s\} : \langle c_i, x\rangle = \gamma_i\}.$$

5. Identification of the faces of convex polyhedra.

Let us now concentrate on the case where C is assumed to be a nonempty polyhedron with representation (4.1) and study some ways to identify the faces of C. In this regard, particular attention is given to the role played by the set $I(x)$ of active constraints at a point $x \in C$.

Given a subset I of $\{1, 2, \cdots, s\}$ we define the set

$$C_I := \{x \in C : \langle c_i, x\rangle = \gamma_i, \quad i \in I\}. \tag{5.1}$$

C_I is always a face of C, in particular it may be the empty face, and if $x \in C_I$, then $I \subset I(x)$. If $I = I(x)$, then

$$C_I = F(x|C)$$

and if $I = \emptyset$, then

$$C_I = C.$$

It can happen that $I \subsetneqq I(x)$ and yet $C_I = F(x|C)$. An example of this phenomenon is given after Definition 5.1. In general, however, if $I_1 \subset I_2 \subset \{1, \cdots, s\}$ and C_{I_2} is not the empty face, then C_{I_1} is also a nonempty face of C with

$$C_{I_2} \subset C_{I_1}.$$

According to Proposition 4.2 the active constraints $I(x)$ at a point x in C are related to the normal cone to C at x via the representation

$$N(x|C) = N_{I(x)} + S \tag{5.2}$$

where $S := \mathrm{Span}\,\{c_i : i = s+1, \cdots, m\}$, and

$$N_I := \left\{ \sum_I u_i c_i : u_i \ge 0, \quad i \in I \right\}. \tag{5.3}$$

Just as it is possible that $C_I = F(x|C)$ for $I \subsetneqq I(x)$, it may be the case that $N(x|C) = N_I + S$ for some $I \subsetneqq I(x)$. Understanding and quantifying these redundancies in the representation of the faces $F(x|C)$ and the normal cones $N(x|C)$ is important to our development, for this reason we introduce the following terminology.

DEFINITION 5.1. Let $C \subset \mathbb{R}^n$ be as in (4.1), $x \in C$, and $I \subset \{1, \cdots, s\}$.

(1) Given $(x, x^*) \in \mathbb{R}^n \times \mathbb{R}^n$ we say that I identifies (x, x^*) as a member of the graph of $N(\cdot|C)$ if $I \subset I(x)$ and $x^* \in N_I + S$.

(2) We say that I identifies $F(x|C)$ if $I \subset I(x)$ and $C_I = F(x|C)$.

(3) We say that I strongly identifies $F(x|C)$ if $I \subset I(x)$ and $N(x|C) = N_I + S$.

Note that if $I \subset I(x)$ strongly identifies $F(x|C)$, then it necessarily identifies $F(x|C)$. To see this, note that C_I is a face of C since $I \subset I(x)$. Also, $N(C_I|C) = N_I + S = N(x|C)$. Hence $C_I = F(x|C)$ by Part (4) of Proposition 4.1. On the other hand, it is possible for the set $I \subset I(x)$ to identify a face $F(x|C)$ but not strongly identify it. For example, consider the set C defined by (4.1) with $s = m = 3$ where $c_1 = \binom{1}{0}$, $c_2 = \binom{0}{1}$, $c_3 = \binom{1}{1}$, and $\gamma_1 = \gamma_2 = \gamma_3 = 0$. Then $I = \{3\}$ identifies $F\binom{0}{0}|C)$, but does not strongly identify it. In the following result we relate these various notions of identification.

THEOREM 5.1. *Let $C \subset \mathbb{R}^n$ be as in* (4.1)*; choose $x \in C$, and let $I \subset I(x)$.*

(1) *The intersection $N_I \cap \operatorname{ri}(N(x|C))$ is nonempty if and only if I identifies the face $F(x|C)$.*

(2) *If I identifies (x, x^*) as a member of the graph of $N(\cdot|C)$ and $x^* \in \operatorname{ri}(N(x|C))$, then I identifies $F(x|C)$.*

(3) *If I strongly identifies $F(x|C)$, then $C = \widetilde{C}$ where*

$$\widetilde{C} := \{x : \langle c_i, x\rangle \le \gamma_i, \quad i \in \widetilde{I}, \langle c_i, x\rangle = \gamma_i, \quad i = s+1, \cdots, m\}$$

with $\widetilde{I} := \{1, 2, \cdots, s\} \backslash (I(x) \backslash I)$.

Proof. (1) ($\Longrightarrow$) Let $x^* \in N_I \cap \operatorname{ri}(N(x|C))$. Then $C_I \subset F(x|C)$, by Part (4) of Proposition 4.1. But $I \subset I(x)$ so that $F(x|C) \subset C_I$. Hence $C_I = F(x|C)$.

($\Longleftarrow$) Let $u_i > 0$ for $i \in I$ and set $x^* := \sum_{i \in I} u_i c_i$. Then x^* is in both $N(x|C)$ and N_I. Hence, by Part (3) of Proposition 4.1, $C_I = F(x|C) \subset E(x^*|C)$. Let $z \in E(x^*|C)$. Then

$$\sum_{i \in I} u_i \langle c_i, z\rangle = \langle x^*, z\rangle = \langle x^*, x\rangle = \sum_{i \in I} u_i \gamma_i$$

since $x \in E(x^*|C)$. Hence $\langle c_i, z\rangle = \gamma_i$ for each $i \in I$. Therefore $z \in C_I = F(x|C)$ so that $E(x^*|C) \subset F(x|C)$. Consequently, $x^* \in \operatorname{ri}(N(x|C))$ by Theorem 4.1.

(2) Since $N_I \cap \operatorname{ri}(N(x|C))$ is nonempty, Part (1) yields the result.

(3) Let $i_0 \in I(x) \backslash I$. Then, since $N(x|C) = N_I + S$, there exist scalars u_i for $i \in I \cup \{s+1, \cdots, m\}$ with $u_i \ge 0$ for $i \in I$ such that

$$c_{i_0} = \sum_I u_i c_i + \sum_{i=s+1}^{n} u_i c_i.$$

Also, since $i_0 \in I(x)$, we have

$$\begin{aligned}\gamma_{i_0} = \langle c_{i_0}, x\rangle &= \sum_I u_i \langle c_i, x\rangle + \sum_{i=s+1}^{m} u_i \langle c_i, x\rangle \\ &= \sum_I u_i \gamma_i + \sum_{i=s+1}^{m} u_i \gamma_i.\end{aligned}$$

Hence, if $y \in \widetilde{C}$, then

$$\begin{aligned}\langle c_{i_0}, y\rangle &= \sum_{i\in I} u_i\langle c_i, y\rangle + \sum_{i=s+1}^{m} u_i\langle c_i, y\rangle \\ &\le \sum_{i\in I} u_i\gamma_i + \sum_{i=s+1}^{m} u_i\gamma_i \\ &= \gamma_{i_0}.\end{aligned}$$

Therefore, $\widetilde{C} \subset C$, hence $\widetilde{C} = C$ since clearly $C \subset \widetilde{C}$. □

6. Constraint identification theory. The identification results that we obtain are similar to those given in [5]. In this regard, we make use of the projection into a convex set. Given a nonempty closed convex set $C \subset \mathbb{R}^n$, recall that the problem

$$\min\{\|y - x\|_2 : x \in C\}$$

has a unique solution in C for each $y \in \mathbb{R}^n$. The solution is called the projection of y into C and is denoted $P_C(y)$. The mapping $P_C : \mathbb{R}^n \to C$ defined in this way has a long and rich history (e.g., see [22]).

The following lemma is the key to most of the results of this section.

LEMMA 6.1. *Let $\overline{x} \in \Omega$ be such that the* MFCQ *is satisfied at $\overline{x}$. Suppose that there are sequences $\{x^i\}, \{r^i\}, \{y^i\}$, and $\{x^{*i}\}$ such that $y^i \in L\Omega(x^i, r^i)$ and $x^{*i} \in N(y^i|L\Omega(x^i, r^i))$ for all $i = 1, 2, \cdots$, with $x^i \to \overline{x}$, $r^i \to 0$, $y^i \to \overline{x}$, and $x^{*i} \to x^*$ for some $x^* \in N(\overline{x}|L\Omega(\overline{x}, 0))$. Then for all i sufficiently large $A(x^i, r^i, y^i)$ identifies $(\overline{x}, x^*)$ as a member of the graph of $N(\cdot|L\Omega(\overline{x}, 0))$. If we further assume that $x^* \in \operatorname{ri}(N(\overline{x}|L\Omega(\overline{x}, 0)))$, then for all i sufficiently large $A(x^i, r^i, y^i)$ identifies the face $F(\overline{x}|L\Omega(\overline{x}, 0))$.*

Proof. We assume that the result does not hold and establish a contradiction. First note that due to the finiteness of the index sets, there is a subsequence $J \subset \mathbb{N}$ such that $A(x^i, r^i, y^i) = A$ for all $i \in J$ where A does not identify $(\overline{x}, x^*)$ as a member of the graph of $N(\cdot|L\Omega(\overline{x}, 0))$. By Lemma 2.1, $A \subset A(\overline{x}, 0, \overline{x})$. Now since $x^{*i} \in N(y^i|L\Omega(x^i, r^i))$ for $i = 1, 2, \cdots$, there exist scalars $u^i_j \ge 0$ for $j \in A$ and $u^i_j \in \mathbb{R}$ for $j = s+1, \cdots, m$ such that

$$x^{*i} = \sum_{A} u^i_j \nabla g_j(x^i) + \sum_{j=s+1}^{m} u^i_j \nabla g_j(x^i). \tag{6.1}$$

Consider the vectors $u^i \in \mathbb{R}^m$ defined by

$$u^i_j = \begin{cases} u^i_j & \text{if } j \in A \cup \{s+1, \cdots, m\} \\ 0 & \text{otherwise}\end{cases}$$

for $j = 1, \cdots, m$ and $i = 1, 2, \cdots$. We claim that the sequence $\{u^i\} \subset \mathbb{R}^m$ is bounded. Indeed, if this were not the case, there would be a subsequence $\widehat{J} \subset J$ such that $\|u^i\| \uparrow^{\widehat{J}} \infty$, and $u^i/\|u^i\| \xrightarrow{\widehat{J}} \overline{u}$ for some $\overline{u} \in \mathbb{R}^m$ with $\|\overline{u}\| = 1$, $\overline{u}_j \ge 0$ for $j \in A$, and $\overline{u}_j = 0$ for $j \ne A \cup \{s+1, \cdots, m\}$. But then, by dividing (6.1) by $\|u^i\|$ and taking the limit over $\widehat{J}$ we get

$$0 = \sum_{A} \overline{u}_j \nabla g_j(\overline{x}) + \sum_{j=s+1}^{m} \overline{u}_j \nabla g_j(\overline{x})$$

while $\|\overline{u}\| = 1$. This contradicts the fact that the MFCQ is satisfied at $\overline{x}$. Hence the sequence $\{u^i\}$ is bounded and with no loss of generality $u^i \to u$ for some $u \in \mathbb{R}^m$ with $u_j \geq 0$ for $j \in A$ and $u_j = 0$ for $j \notin A \cup \{s+1, \cdots, m\}$. But then

$$x^* = \sum_A u_j \nabla g_j(\overline{x}) + \sum_{j=s+1}^{m} u_j \nabla g_j(\overline{x})$$

so that A identifies $(\overline{x}, x^*)$ as a member of the graph of $N(\cdot|L\Omega(\overline{x}, 0))$. This contradiction establishes the result.

The last statement of the lemma now follows immediately from Part (2) of Theorem 5.1 □

THEOREM 6.1. *Let $\overline{x} \in \Omega$ be a stationary point of* NLP *at which the* MFCQ *is satisfied. Suppose that $\{(x^i, r^i, y^i)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ is a sequence such that $y^i \in L\Omega(x^i, r^i)$ for all $i = 1, 2, \cdots$, and $(x^i, r^i, y^i) \to (\overline{x}, 0, \overline{x})$. Then $A(x^i, r^i, y^i)$ identifies $(\overline{x}, -\nabla f(\overline{x}))$ as a member of the graph of $N(\cdot|L\Omega(\overline{x}, 0))$ for all i sufficiently large if and only if*

$$P_{T^i}(-\nabla f(x^i)) \to 0 \tag{6.2}$$

where $T^i := T(y^i|L\Omega(x^i, r^i))$ for each $i = 1, 2, \cdots$.

Remark. Theorem 6.1 is quite similar to [5, Thm. 3.4] wherein the finite attainment of the active constraint set was characterized in terms of the quantity $P_{T(x|\Omega)}(-\nabla f(x))$ referred to as the projected gradient.

Proof. Suppose that $A(x^i, r^i, y^i)$ identifies $(\overline{x}, -\nabla f(\overline{x}))$ as a member of the graph of $N(\cdot|L\Omega(\overline{x}, 0))$ for all i sufficiently large. Due to the finiteness of the index set, there are at most a finite number of index sets $A \subset \{1, \cdots, s\}$ which identify $(\overline{x}, -\nabla f(\overline{x}))$ as a member of the graph of $N(\cdot|L\Omega(\overline{x}, 0))$. Suppose there are q of them which we denote by $A_k = 1, \cdots, q$. Then, for each $k = 1, \cdots, q$, there are vectors $u^k \in \mathbb{R}^m$ such that

$$\begin{aligned} -\nabla f(\overline{x}) &= g'(\overline{x})^T u^k \\ 0 &= u_j^k \quad \text{for } j \in \{1, \cdots, s\} \backslash A_k, \end{aligned}$$

and

$$0 \leq u_j^k \quad \text{for } j \in A_k$$

where $g'(\overline{x}) := [\nabla g_1(\overline{x}), \cdots, \nabla g_m(\overline{x})]^T$. Hence, for all i sufficiently large

$$\begin{aligned} \|P_{T_i}(-\nabla f(x^i))\| &\leq \min \left\{ \|\nabla f(x^i) + g'(x^i)^T u\| \;\middle|\; \begin{array}{l} u_j \geq 0 \;\; \text{for } j \in A(x^i, r^i, y^i) \\ u_j = 0 \;\; \text{for } j \in \{1, \cdots, s\} \backslash A(x^i, r^i, y^i) \end{array} \right\} \\ &\leq \max\{\|\nabla f(x^i) + g'(x^i)^T u^k\| : k = 1, \cdots, q\}. \end{aligned}$$

Thus (6.2) follows by continuity.

Conversely, suppose that (6.2) occurs and define $x^{*i} := P_{N^i}(-\nabla f(x^i))$ where $N^i := N(y^i|L\Omega(x^i, r^i))$ for each $i = 1, 2, \cdots$. Then the Moreau decomposition of $-\nabla f(x^i)$ is

$$-\nabla f(x^i) = x^{*i} + P_{T^i}(-\nabla f(x^i)). \tag{6.3}$$

From (6.2) and (6.3) we get that $x^{*i} \to -\nabla f(\overline{x}) \in N(\overline{x}|L\Omega(\overline{x},0))$. Hence Lemma 6.1 and Theorem 3.1 apply yielding the result. □

COROLLARY 6.1.1. *Let the assumptions of Theorem* 6.1 *hold and further suppose that* $\overline{x}$ *is a nondegenerate stationary point for* NLP. *Then* $A(x^i, r^i, y^i)$ *identifies the face* $F(\overline{x}|L\Omega(\overline{x},0))$ *for all* i *sufficiently large if and only if* $P_{T^i}(-\nabla f(x^i)) \to 0$.

Proof. The result follows immediately from the theorem and the last sentence of Lemma 6.1 with $x^{*i} = P_{N^i}(-\nabla f(x^i))$ where $N^i := N(y^i|L\Omega(x^i,r^i))$. □

In practice one of the primary applications of finite identification results is toward the establishment of local rates of convergence. In this context we usually require knowledge of the active set of indices at the solution, $A(\overline{x},0,\overline{x})$. To ensure this identification property the LI condition (linear independence condition) is usually invoked.

DEFINITION 6.1. Let $\overline{x} \in \Omega$. We say that the LI condition is satisfied at $\overline{x}$ if the vectors

$$\{\nabla g_i(\overline{x}) : i \in A(\overline{x},0,\overline{x}) \cup \{s+1,\cdots,m\}\}$$

are linearly independent.

The LI condition is a convenient tool for this purpose since it is a sufficient condition under which every representation of an element of $\mathrm{ri}\,(N(\overline{x}|L\Omega(x,0)))$ must explicitly employ each of the active constraints.

LEMMA 6.2. *Suppose that the* LI *condition is satisfied at the point* $\overline{x} \in \Omega$. *Then the point*

$$x^* = \sum_{i\in A(x,0,x)} u_i \nabla g_i(\overline{x}) + \sum_{i=s+1}^{m} u_i \nabla g_i(\overline{x})$$

is an element of $\mathrm{ri}\,(N(\overline{x}|L\Omega(x,0))$ *only if* $u_i > 0$ *for each* $i \in A(\overline{x},0,\overline{x})$.

The proof of this lemma is straightforward and so is left as an exercise for the reader. This observation and Theorem 6.1 combine to yield our main result on constraint identification.

THEOREM 6.2. *Suppose* $\overline{x} \in \Omega$ *is a nondegenerate stationary point of* NLP *at which the* LI *condition is satisfied. If* $\{(x^i,r^i,y^i)\} \subset \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ *is a sequence such that* $y^i \in L\Omega(x^i,r^i)$ *for all* $i = 1,2,\cdots$, *and* $(x^i,r^i,y^i) \to (\overline{x},0,\overline{x})$, *then*

$$A(x^i,r^i,y^i) = A(\overline{x},0,\overline{x})$$

for all i *sufficiently large if and only if* (6.2) *occurs.*

Proof. It is well known and easily established that the LI condition implies the MFCQ [15], hence the assumptions of Theorem 6.1 are satisfied. Thus we need only show that an index set $A \subset A(\overline{x},0,\overline{x})$ identifies $(\overline{x}, -\nabla f(\overline{x}))$ as a member of the graph of $N(\cdot|L\Omega(\overline{x},0))$ if and only if $A = A(\overline{x},0,\overline{x})$. If $A = A(\overline{x},0,\overline{x})$, then A trivially identifies $(\overline{x}, -\nabla f(\overline{x}))$ as a member of the graph of $N(\cdot|L\Omega(\overline{x},0))$. The converse implication follows immediately from Lemma 6.2 and Part (1) of Definition 5.1. □

A natural question at this point is to ask whether we really require the LI condition to establish a result similar to Theorem 6.2. Since the usefulness of the LI condition is derived from its application in Lemma 6.2, one way to view this question is to determine whether the LI condition is a necessary condition under which every representation of an element of $\mathrm{ri}\,(N(\overline{x}|L\Omega(\overline{x},0)))$ must explicitly employ each of the active constraints.

LEMMA 6.3. *Let* $\overline{x} \in \Omega$. *If any point of the form*

$$x^* = \sum_{i\in A(x,0,x)} u_i \nabla g_i(\overline{x}) + \sum_{i=s+1}^{m} u_i \nabla g(\overline{x})$$

is an element of $\mathrm{ri}\,(N(\overline{x}|L\Omega(\overline{x},0)))$ *only if* $u_i > 0$ *for each* $i \in A(\overline{x},0,\overline{x})$, *then the vectors* $\{\nabla g_i(\overline{x}) : i \in A(\overline{x},0,\overline{x})\}$ *are linearly independent.*

Proof. Let $\{v_i : i \in A(\overline{x},0,\overline{x})\} \subset \mathbb{R}$ be such that $0 = \sum_{i\in A(\overline{x},0,\overline{x})} v_i \nabla g_i(\overline{x})$ and consider $u \in \mathbb{R}^m$ such that $u_i > 0$ for each $i \in A(\overline{x},0,\overline{x})$. Then, by (3.7),

$$x^* = \sum_{i\in A(\overline{x},0,\overline{x})} (u_i + tv_i)\nabla g_i(\overline{x}) \in \mathrm{ri}\,(N(\overline{x}|L\Omega(\overline{x},0)))$$

for every $t \in \mathbb{R}$. If the gradients $\{\nabla g_i(\overline{x}) : i \in A(\overline{x},0,\overline{x})\}$ are not linearly independent then with no loss of generality the set of indices $\{i : v_i < 0, i \in A(\overline{x},0,\overline{x})\} =: I_-$ is nonempty. Now choose $t > 0$ to make one of the $u_i + tv_i$ zero while the others remain nonnegative. But then, by hypothesis, $x^* \notin \mathrm{ri}\ (N(\overline{x}|L\Omega(\overline{x},0)))$ and this is a contradiction. □

Now, given that the MFCQ implies that the constraint gradients $\{\nabla g_i(\overline{x}) : i = s+1,\cdots,m\}$ are linearly independent, Lemmas 6.2 and 6.3 and Theorem 6.2 seem to imply that for general sequences $(x^i, r^i, y^i) \to (\overline{x},0,\overline{x})$ the LI condition is in a sense required if we wish to establish a general finite identification result for the active constraints. This leads us to the deeper question of whether knowledge of the active constraint set is really required. For example, would it suffice to only identify a set of constraints that strongly identify the optimal face of $F(\overline{x}|L\Omega(\overline{x},0))$? Part (3) of Theorem 5.1 suggests that this may be the case. However, in order to use Corollary 6.1.1 to establish that $F(\overline{x}|L\Omega(\overline{x},0))$ is strongly identified after a finite number of iterations without further assumptions on the structure of the sequence $\{(x^i, r^i, y^i)\}$ we would need to know that every set of indices that identifies the face $F(\overline{x}|L\Omega(\overline{x},0))$ also strongly identifies it. Unfortunately, it can be shown that this condition on the face $F(\overline{x}|L\Omega(\overline{x},0))$ is virtually equivalent to the LI condition.

THEOREM 6.3. *Let the* MFCQ *be satisfied at* $\overline{x} \in \Omega$. *Then every index set* $A \subset A(\overline{x},0,\overline{x})$ *that identifies the face* $F(\overline{x}|L\Omega(\overline{x},0))$ *also strongly identifies the face* $F(\overline{x}|L\Omega(\overline{x},0))$ *if and only if there is a partition of the set of indices* $A(\overline{x},0,\overline{x})$, *say* $\{E_k : k = 1,\cdots,\ell\}$, *such that*

(a) *for each* $k = 1,\cdots,\ell$ *and each pair* i,j *of elements of* E_k *there is a* $\lambda > 0$ *for which* $P_{S^\perp}(\nabla g_i(\overline{x})) = \lambda P_{S^\perp}(\nabla g_j(\overline{x}))$ *where* $S := \mathrm{Span}\,\{\nabla g_i(\overline{x}) : i = s+1,\cdots,m\}$, *and*

(b) *for each selection from the partition,* $\{i_k : k = 1,\cdots,\ell\}$ *with* $i_k \in E_k$, *the vectors* $\{\nabla g_{i_k}(\overline{x}) : k = 1,\cdots,\ell\} \cup \{\nabla g_i(\overline{x}) : i = s+1,\cdots,m\}$ *are linearly independent.*

The proof of Theorem 6.3 is omitted here since it is far too lengthy and contributes little insight into the problem at issue. This proof is provided in Appendix B.

Theorem 6.3 and Lemmas 6.2 and 6.3 seem to indicate that if we wish to employ the techniques of this paper to demonstrate the finite identification of either the set of active constraints or a set of indices that strongly identifies the optimal face of $L\Omega(\overline{x},0)$, then the LI condition is in essence required to hold at $\overline{x} \in \Omega$. From our perspective this is a negative result when inequality constraints are present. Consequently, more effort and perhaps more sophisticated techniques need to be applied to this problem.

7. Application to existing algorithms. In this section we consider two algorithmic frameworks for constrained optimization. The first framework is based upon the SQP algorithm of Wilson [19], Han [10], and Powell [13], and the second is based upon the QL (or S_{ℓ_1} QP) algorithm of Fletcher [8],[9]. In each case we will assume that the algorithm generates a triple (x^i, r^i, y^i) with $y^i \in L\Omega(x^i, r^i)$, $r^i \to 0$, and $\|x^i - y^i\| \to 0$, and show that

$$P_{T(y^i|L\Omega(x^i,r^i))}(-\nabla f(x^i)) \to 0. \tag{7.1}$$

The results are obtained without assuming that the x^i's actually converge. Nonetheless, the two assumptions $r^i \to 0$ and $\|x^i - y^i\| \to 0$ do imply that

$$\operatorname{dist}(g(x^i)|K) \to 0$$

where $K := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$. Once (7.1) is established, the results of the previous sections, in particular Theorem 6.1, Corollary 6.11, and Theorem 6.2, can be applied to yield various statements concerning the identification of active constraints for these algorithms.

7.1. The SQP Algorithm.

(0) Set $\Pi_{\text{SQP}} := \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}_+ \times (\mathbb{R}_+\backslash\{0\}) \times \mathbb{R}^{n\times n}$ and choose $(x^0, w^0, r^0, \kappa_0, \delta_0, H_0) \in \Pi_{\text{SQP}}$ so that the program Q^0_{SQP} (defined in step 1) is feasible.

(1) If x^i is a stationary point of NLP stop; otherwise, let y^i be a stationary point for the program

$$\begin{aligned} Q^i_{\text{SQP}} : &\min(\nabla f(x^i) + w^i)^T(y - x^i) + \tfrac{1}{2}(y - x^i)^T H_i(y - x^i) \\ &\text{subject to } y \in x^i + \delta_i\mathbb{B} \text{ and} \\ &g(x^i) + g'(x^i)(y - x^i) - r^i \in K + \kappa_i\mathbb{B} \end{aligned}$$

where $K := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$.

(2) Compute the updates

$$(x^{i+1}, w^{i+1}, r^{i+1}, \kappa_{i+1}, \delta_{i+1}, H_{i+1}) \in \Pi_{\text{SQP}}$$

so that the convex program Q^{i+1}_{SQP} is feasible. Set $i = i + 1$ and return to Step 1.

Remarks. The parameters w and r can be used to represent inexact solution techniques [21], the parameter κ can be used to represent modified SQP Newton iterates [3], [4], [18], and δ can be used to represent a possible trust region variation [3], [4].

The following technical lemma will greatly facilitate our discussion.

LEMMA 7.1. *Let $(x, r, y) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ be such that*

$$y \in L\Omega_\kappa(x, r) := \{z : g(x) + g'(x)(z - x) - r \in K + \kappa\mathbb{B}\}$$

for some $\kappa > 0$, where $K := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$.

If we define $\hat{r} \in \mathbb{R}^m$ componentwise by

$$\hat{r}_j := (g_j(x) + \nabla g_j(x)^T(y - x))_+ \quad \textit{for } j = 1, \cdots, s$$

where $\xi_+ := \max\{0, \xi\}$ for all $\xi \in \mathbb{R}$, and

$$\hat{r}_j := g_j(x) + \nabla g_j(x)^T(y - x) \quad \textit{for } j = s + 1, \cdots, m,$$

then

$$\|P_{T(y|L\Omega(x,\hat{r}))}(-\nabla f(x))\| \le \operatorname{dist}[-\nabla f(x)|N(y|L\Omega_\kappa(x, r))].$$

Proof. First observe that

$$L\Omega(x,\hat{r}) \subset L\Omega_\kappa(x,r).$$

Indeed, since $\operatorname{dist}[g(x)+g'(x)(y-x)-r|K] = \operatorname{dist}[\hat{r}-r|K]$ and $\operatorname{dist}[g(x)+g'(x)(z-x)-\hat{r}|K]=0$ for every $z \in L\Omega(x,\hat{r})$, we have that

$$\begin{aligned}
&\operatorname{dist}[g(x)+g'(x)(z-x)-r|K] \\
&\quad\le \operatorname{dist}[g(x)+g'(x)(z-x)-\hat{r}|K] + \operatorname{dist}[\hat{r}-r|K] \\
&\quad= \operatorname{dist}[g(x)+g'(x)(y-x)-r|K] \\
&\quad\le \kappa,
\end{aligned}$$

for every $z \in L\Omega(x,\hat{r})$, which implies that $z \in L\Omega_\kappa(x,r)$.

Now since $L\Omega(x,\hat{r}) \subset L\Omega_\kappa(x,r)$ where both of these sets are convex and contain the point y, we have

$$T(y|L\Omega(x,\hat{r})) \subset T(y|L\Omega_\kappa(x,r)).$$

Hence

$$N(y|L\Omega_\kappa(x,r)) \subset N(y|L\Omega(x,\hat{r})).$$

Consequently,

$$\begin{aligned}
\|P_{T(y|L\Omega(x,\hat{r}))}(-\nabla f(x))\| &= \operatorname{dist}[-\nabla f(x)|N(y|L\Omega(x,\hat{r}))] \\
&\le \operatorname{dist}[-\nabla f(x)|N(y|L\Omega_\kappa(x,r))]. \qquad \square
\end{aligned}$$

We now establish (7.1) for the SQP Algorithm 7.1 by demonstrating that

$$\operatorname{dist}[-\nabla f(x^i)|N(y^i|L\Omega_{\kappa_i}(x^i,r^i))] \to 0.$$

THEOREM 7.1. *Let $\{(x^i,w^i,r^i,\kappa_i,\delta_i,H_i)\}$ be a sequence generated by the* SQP *Algorithm* 7.1 *and suppose that $\|y^i-x^i\| \to 0$ with $\|y^i-x^i\| < \delta_i$ for all i sufficiently large, $w^i \to 0$, and $\{H_i\}$ is bounded, then*

$$P_{T(y^i|L\Omega(x^i,\hat{r}^i))}(-\nabla f(x^i)) \to 0$$

where

$$\hat{r}^i_j := (g_j(x^i)+\nabla g_j(x^i)^T(y^i-x^i))_+ \quad \textit{for } j=1,\cdots,s$$

$$\hat{r}^i_j := (g_j(x^i)+\nabla g_j(x^i)^T(y^i-x^i)) \quad \textit{for } j=s+1,\cdots,m.$$

Proof. By Lemma 7.1 we need only show that

$$\operatorname{dist}[-\nabla f(x^i)|N(y^i|L\Omega_{\kappa_i}(x^i,r^i))] \to 0.$$

The hypotheses and Step (1) of the algorithm imply that

$$-[\nabla f(x^i)+w^i+H_i(y^i-x^i)] \in N(y^i|L\Omega_{\kappa_i}(x^i,r^i))$$

for all i sufficiently large. Hence

$$\operatorname{dist}[-\nabla f(x^i)|N(y^i|L\Omega_{\kappa_i}(x^i,r^i))] \le \|w^i\| + \|H_i\|\ \|y^i-x^i\|,$$

for all i sufficiently large, whereby the result is established. $\square$

Observe that we require the eventual inactivity of the constraint $\|y^i - x^i\| \le \delta_i$. The reasonableness of this hypothesis depends upon algorithmic details and second-order information at a cluster point of $\{x_i\}$. For example, if $\|y^i - x^i\| \to 0$, the constraint $\|y^i - x^i\| \le \delta_i$ is eventually inactive for the algorithms described in [3], [4].

Let us now consider a model algorithm due to Fletcher [8], [9] which we refer to as the QL algorithm. This proceedure is motivated by the fact that a point $\overline{x} \in \Omega$ is a Kuhn–Tucker point for NLP if and only if there is an $\overline{\alpha} > 0$ such that $\overline{x}$ is a stationary point for the exact penalty function

$$P_\alpha(x) := f(x) + \alpha \text{dist}\,[g(x)|K]$$

for all $\alpha \ge \overline{\alpha}$ where again $K := \mathbb{R}^s_- \times \{0\}_{\mathbb{R}^{m-s}}$ and the distance function dist$\,(g(x)|K)$ may be defined using any norm on $\mathbb{R}^m$ (e.g., see [11]).

A point $\overline{x} \in \mathbb{R}^n$ is a stationary point of $P_\alpha(x)$ if $0 \in \partial P_\alpha(\overline{x})$ where $\partial P_\alpha(x)$ is the Clarke subdifferential of P_α at x (e.g., see [6]).

7.2. The QL Algorithm.

(0) Set $\Pi_{\text{QL}} := \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R}_+\backslash\{0\}) \times \mathbb{R}_+ \times \mathbb{R}^{n\times n}$ and choose

$$(x^0, w^0, r^0, \delta_0, \alpha_0, H_0) \in \Pi_{\text{QL}}$$

with $\alpha_0 \ge \overline{\alpha}$ for some $\overline{\alpha} > 0$.

(1) If x^i is a stationary point of the exact penalty function P_{α_i}, stop; otherwise, let y^i be a stationary point of the convex program

$$Q^i_{QL} : \min\{\varphi_i(y) : y \in x^i + \delta_i \mathbb{B}\}$$

where

$$\varphi_i(y) := (\nabla f(x^i) + w^i)^T (y - x^i) + \tfrac{1}{2}(y - x^i)^T H_i (y - x^i) + \alpha_i \text{dist}\,[g(x^i) + g'(x^i)(y - x^i) - r^i|K].$$

(2) Compute the updates

$$(x^{i+1}, w^{i+1}, r^{i+1}, \delta_{i+1}, \alpha_{i+1}, H_{i+1}) \in \Pi_{QL}$$

with $\alpha_{i+1} \ge \overline{\alpha}$. Set $i = i + 1$ and return to Step (1).

Remark. The parameters w^i, r^i, and δ_i play roles similar to those that they play in the SQP algorithm.

THEOREM 7.2. *Let $\{(x^i, w^i, r^i, \delta_i, H_i)\}$ be a sequence generated by the* QL *algorithm and suppose that $\|y^i - x^i\| \to 0$, $w^i \to 0$, the sequence $\{H_i\}$ is bounded, and for all i sufficiently large $\|y^i - x^i\| < \delta_i$. Then*

$$P_{T(y^i|L\Omega(x^i,\widehat{r}^i))}(-\nabla f(x^i)) \to 0$$

where

$$\widehat{r}^{\,i}_j := (g_j(x^i) + \nabla g_j(x^i)^T (y^i - x^i))_+ \quad \textit{for } i = 1, \cdots, s,$$

$$\widehat{r}^{\,i}_j := g_j(x^i) + \nabla g_j(x^i)^T (y^i - x^i) \quad \textit{for } i = s+1, \cdots, m.$$

Proof. Again by Lemma 7.1 we need only show that

$$\text{dist}\,[-\nabla f(x^i)|N(y^i|L\Omega_{K_i}(x^i, r^i))] \to 0$$

where

$$\kappa_i := \text{dist}\,[g(x^i) + g'(x^i)(y^i - x^i) - r^i | K].$$

is chosen to ensure that $y^i \in L\Omega_{\kappa_i}(x^i, r^i)$.

Since y^i is a stationary point for Q^i_{QL} and $\|y^i - x^i\| < \delta_i$ for i sufficiently large, we know that

$$-[\nabla f(x^i) + w^i + H_i(y^i - x^i)]$$

is an element of the set

$$\alpha_i g'(x^i)^T \partial \text{dist}\,[\cdot|K](g(x^i) + g'(x^i)(y^i - x^i) - r^i)$$

for all i sufficiently large by [6, Chap. 2]. Also, by [2, §2], we know that

$$\partial \text{dist}\,[\cdot|K](z) := \begin{cases} (\text{bdry}\mathbb{B}^0) \cap N(z|K + \text{dist}\,(z|K)\mathbb{B}) & \text{if } z \notin K \\ \mathbb{B}^0 \cap N(z|K) & \text{if } z \in K. \end{cases}$$

where bdry $\mathbb{B}^0$ is the boundary of $\mathbb{B}^0$. Hence we have that

$$-[\nabla f(x^i) + w^i + H_i(y^i - x^i)]$$

is an element of

$$N_i := g'(x^i)^T N[g(x^i) + g'(x^i)^T(y^i - x^i) - r^i | K + \kappa_i \mathbb{B}]$$

for all i sufficiently large. But by [16, Thm. 23.9],

$$N_i \subset N(y^i | L\Omega_{\kappa_i}(x^i, r^i)).$$

Hence

$$\begin{aligned} &\text{dist}\,[-\nabla f(x^i) | N(y^i | L\Omega_{\kappa_i}(x^i, r^i))] \\ &\leq \text{dist}\,[-\nabla f(x^i) | N_i] \\ &\leq \|w^i\| + \|H_i\|\ \|y^i - x^i\| \end{aligned}$$

for all i sufficiently large, whereby the result is established. □

Again note the use of the hypothesis $\|y^i - x^i\| < \delta_i$ for all i sufficiently large. In the context of a trust region algorithm the reasonableness of this hypothesis relies upon the updating strategy employed in step (2) of the algorithm and the second-order behavior at a cluster point of the sequence $\{x^i\}$. Furthermore, given such information it may be possible to demonstrate that (7.1) occurs even though the constraint $\|y^i - x^i\| \leq \delta_i$ remains active. However, such results are not within the scope of this paper.

Appendix A. The primary purpose of this appendix is to provide the background necessary to prove Theorem 4.1. This theorem appears as Corollary A.1.1. It follows easily from the main theorem of the Appendix which characterizes when a face of a nonempty closed convex set is exposed and identifies those vectors that expose it. The main result is a consequence of the following technical lemma.

LEMMA A.1. *Let $K \subset \mathbb{R}^n$ be a nonempty closed convex cone and let $x^* \in [\text{lin}\,(K)]^\perp$. Then $x^* \in \text{ri}\,(K^0)$ if and only if there is an $\varepsilon > 0$ such that $\langle x^*, y \rangle \leq -\varepsilon \|y\|$ for all $y \in K \cap [\text{lin}\,(K)]^\perp$ where* $\text{lin}\,(K) := K \cap (-K)$.

Proof. From [5, Lem. 2.7] we know that aff $(K^0) = [\text{lin}\,(K)]^\perp$. Hence, if $x^* \in$ ri (K^0), then there is an $\varepsilon > 0$ such that

$$x^* + y^* \in \text{ri}(K^0)$$

for all $y^* \in [\text{lin}(K)]^\perp \cap (\varepsilon \mathbb{B}^0)$. Consequently,

$$0 \geq \langle x^* + y^*, y\rangle = \langle x^*, y\rangle + \langle y^*, y\rangle \tag{A.1}$$

for all $y^* \in [\text{lin}(K)]^\perp \cap (\varepsilon \mathbb{B}^0)$ and $y \in K$. Now if $y \in K \cap [\text{lin}(K)]^\perp$ then y^* can be chosen in (1) so that $\langle x^*, y\rangle \leq -\varepsilon\|y\|$.

Conversely, let $\varepsilon > 0$ be such that $\langle x^*, y\rangle \leq -\varepsilon\|y\|$ for all $y \in K \cap (\text{lin}(K)]^\perp$. From [16, p. 65], for each $y \in K$ there exist $y^1 \in \text{lin}(K)$ and $y^2 \in K \cap [\text{lin}(K)]^\perp$ such that $y = y^1 + y^2$. Thus if $y \in K$ and $y^* \in [\text{lin}(K)]^\perp \cap (\varepsilon \mathbb{B}^0) = [\text{aff}(K^0)] \cap (\varepsilon \mathbb{B}^0)$ (by [5, Lem. 2.7]), then

$$\begin{aligned}\langle x^* + y^*, y\rangle &= \langle x^*, y^2\rangle + \langle y^*, y^2\rangle \\ &\leq -\varepsilon\|y^2\| + \varepsilon\|y^2\| \\ &= 0.\end{aligned}$$

Therefore, $x^* + [(\varepsilon \mathbb{B}^0) \cap \text{aff}(K^0)] \subset K^0$ so that $x^* \in \text{ri}(K^0)$. □

THEOREM A.1. *Let C be a nonempty closed convex subset of $\mathbb{R}^n$ and let $x \in C$.*

(1) *$x^* \in \text{ri}\ (N(x|C))$ if and only if $E(x^*|C) = C \cap [x + \text{lin}(T(x|C))]$.*

(2) *$F(x|C) = [x + \text{lin}(T(x|C))] \cap C$ if and only if there is an $x^* \in \mathbb{R}^n$ such that $E(x^*|C) = F(x|C)$.*

(3) *If $F(x|C) = C \cap [x + \text{lin}(T(x|C))]$, then $x^* \in \text{ri}\, N(x|C)$ if and only if $E(x^*|C) = F(x|C)$.*

Proof. (1) By Lemma A.1 we know that $x^* \in \text{ri}(N(x|C))$ if and only if there is an $\varepsilon > 0$ such that $\langle x^*, y\rangle \leq -\varepsilon\|y\|$ for all $y \in T(x|C) \cap [\text{lin}(T(x|C))]^\perp$. Also, for each $z \in C$, $z - x \in T(x|C)$ and, by [16, p. 65], there is a $z^1(x) \in \text{lin}(T(x|C))$ and a $z^2(x) \in \underline{T(x|C) \cap [\text{lin}(T(x|C))]^\perp}$ such that $z - x = z^1(x) + z^2(x)$. Hence, since $T(x|C) = \cup_{\lambda>0}\lambda^{-1}[C - x]$, these observations imply that $x^* \in \text{ri}(N(x|C))$ if and only if there is an $\varepsilon > 0$ such that

$$\begin{aligned}\langle x^*, z\rangle &= \langle x^*, x\rangle + \langle x^*, z - x\rangle \\ &= \langle x^*, x\rangle + \langle x^*, z^2(x)\rangle \\ &\leq \langle x^*, x\rangle - \varepsilon\|z^2(x)\|\end{aligned}$$

for each $z \in C$ (observe that $\langle x^*, z^1(x)\rangle = 0$ since, by [5, Lem. 2.7], $\text{aff}(N(x|C)) = [\text{lin}(T(x|C))]^\perp$). Therefore, $x^* \in \text{ri}(N(x|C))$ if and only if

$$\begin{aligned}E(x^*|C) &= \{z \in C : z^2(x) = 0\} \\ &= C \cap [x + \text{lin}(T(x|C))].\end{aligned}$$

(2) ($\Longrightarrow$) By (1) any $x^* \in \text{ri}(N(x|C))$ will do.

($\Longleftarrow$) By [5, §2],

$$F(x|C) = \text{aff}(F(x|C)) \cap C \subset x + \text{lin}(T(x|C)). \tag{A.2}$$

Moreover,

$$x + \text{lin}(T(x|C)) \subset E(x^*|C). \tag{A.3}$$

To see this note that $\langle x^*, z\rangle \le 0$ for all $z \in T(x|C)$. Hence $\langle x^*, z\rangle = 0$ for all $z \in \operatorname{lin}(T(x|C))$. Therefore, $\langle x^*, z\rangle = \langle x^*, x\rangle$ for all $z \in x + \operatorname{lin}(T(x|C))$ yielding (A.3). Now (A.2), (A.3), and the hypothesis $E(x^*|C) = F(x|C)$ imply

$$F(x|C) = (x + \operatorname{lin}(T(x|C))) \cap C.$$

(3) This statement follows immediately from Parts (1) and (2). □

COROLLARY A.1.1. *Let $C \in \mathbb{R}^n$ be a nonempty polyhedron and let $x \in C$. Then $E(x^*|C) = F(x|C)$ if and only if $x^* \in \operatorname{ri}(N(x|C))$.*

Proof. By [5, Th. 2.6] we know that $\operatorname{aff}(F(x|C)) = x + \operatorname{lin}(T(x|C))$. Hence the result follows from Part (3) of Theorem A.1. □

Appendix B. This appendix provides a proof of Theorem 6.3. Some preliminary results are first established.

DEFINITION B.1. Let $K \subset \mathbb{R}^n$ be a nonempty closed convex cone. A direction $d \in \mathbb{R}^n\backslash\{0\}$ is said to be an extreme direction of K if the ray $\{\lambda d : \lambda \ge 0\}$ is a face of K.

LEMMA B.1. *Let $K \subset \mathbb{R}^n$ be a nonempty polyhedral cone with representation*

$$K := \left\{ \sum_{i=1}^{s} u_i a_i + \sum_{i=s+1}^{m} u_i a_i : u_i \ge 0,\ i = 1, \cdots, s \right\}$$

where $\{a_i : i = 1, \cdots, m\} \subset \mathbb{R}^n$.

(1) $\operatorname{lin}(K) = \operatorname{Span}\{a_i : a_i \in \operatorname{lin}(K)\}$.

(2) *Every extreme direction of K is of the form λa_i for some $\lambda > 0$ and $i \in \{1, \cdots, s\}$.*

(3) *Let $I \subset \{1, \cdots, s\}$ be such that $i \in I$ if and only if a_i is an extreme direction for K. Then*

$$K = \left\{ \sum_{I} u_i a_i : u_i \ge 0, i \in I \right\} + \operatorname{lin}(K). \qquad \square$$

The proof of the above lemma is omitted as it easily follows from the results in [16, §18]. The main technical lemma required for the proof of Theorem 6.3 now follows.

LEMMA B.2. *Let $C \subset \mathbb{R}^n$ be a nonempty polyhedral convex set with representation (4.1). For each $i \in \{1, \cdots, m\}$ let $\bar{c}_i$ be the orthogonal projection of c_i onto $[\operatorname{lin}(N(x|C))]^\perp$. Let $\{d_i : i = 1, \cdots, \ell\}$ be distinct extreme directions for $N(x|C) \cap [\operatorname{lin}(N(x|C))]^\perp$ such that*

$$\operatorname{cone}\{d_i : i = 1, \cdots, \ell\} = N(x|C) \cap [\operatorname{lin}(N(x|C))]^\perp$$

and set

$$E_j := \{i \in \{1, \cdots, s\} : \bar{c}_i = \lambda d_j \text{ for some } \lambda > 0\}.$$

Then the face $F(x|C)$ is such that the set of indices $I \subset \{1, \cdots, m\}$ that identify $F(x|C)$ equals the set of indices that strongly identify $F(x|C)$ if and only if

(a) $\operatorname{lin}[N(x|C)] = S$,

(b) *the extreme directions $\{d_j : j = 1, \cdots, \ell\}$ are linearly independent, and*

(c) $\bigcup_{j=1}^{\ell} E_j = \{i : c_i \notin S\}$,

where $S := \text{Span}\{c_i : i = s+1, \cdots, m\}$.

Proof. ($\Longrightarrow$) For each $j = 1, \cdots, \ell$ choose $i_j \in E_j$ and set $I = \{i_j : j = 1, \cdots, \ell\}$. Let $i_0 \in \{1, \cdots, s\}$ be such that $c_{i_0} \in \text{lin}(N(x|C))$. Set

$$x^* = c_{i_0} + \sum_I u_i c_i$$

where $u_i > 0$, $i \in I$. Then $x^* \in \text{ri}(N(x|C))$ so that $\widetilde{I} = \{i_0\} \cup I$ identifies $F(x|C)$, by Part (1) of Theorem 5.1. Thus, by hypothesis, $\widetilde{I}$ strongly identifies $F(x|C)$. Therefore $\text{lin}(N(x|C)) = \text{Span}\{c_{i_0}, c_i : i = s+1, \cdots, m\}$. Now $-c_{i_0} \in \text{lin}(N(x|C))$. Hence there are scalars v_i and $\lambda \geq 0$ such that $-c_{i_0} = \lambda c_{i_0} + \sum_{i=s+1}^{n} v_i c_i$. Consequently, $c_{i_0} \in S$, whereby $\text{lin}(N(x|C)) = S$.

We now establish the linear independence of the vectors $\{c_i : i \in I\}$. If $\ell = 0$, then the result follows trivially, thus we assume that $\ell \geq 1$. Let J be any nonempty subset of I and consider the cone $N_{I \setminus J}$. If

$$N_{I\setminus J} \cap \text{ri}(N(x|C)) \neq 0,$$

then $C_{I\setminus J} = F(x|C)$ by Part (1) of Theorem 5.1. In this case $I\setminus J$ strongly identifies $F(x|C)$. But

$$N(x|C) \neq N_{I\setminus J} + S$$

as $\bar{c}_i \notin N_{I\setminus J} + S$ for $i \in J$. Therefore,

$$N_{I\setminus J} \subset \text{rbdry}(N(x|C)), \tag{B.1}$$

where rbdry $(N(x|C))$ is the relative boundary of $N(x|C)$.

Next let $v_i \in \mathbb{R}$ for $i \in I$ be such that $0 = \sum_I v_i \bar{c}_i$. Choose $u_i > 0, i \neq I$ and set $x^* = \sum_I u_i \bar{c}_i$. Now $x^* \in \text{ri}(N(x|C))$ and

$$x^* = \sum_I (u_i + \mu v_i)\bar{c}_i$$

for all $\mu \in \mathbb{R}$. Assuming $I_- := \{i : v_i < 0\}$ is nonempty, set

$$t := \min\left\{\frac{u_i}{(-v_i)} : i \in I_-\right\},$$

and let $i_0 \in I_-$ be such that $t = u_{i_0}/(-v_{i_0})$. Then

$$x^* = \sum_{I\setminus\{i_0\}} (u_i + tv_i)c_i$$

with $(u_i + tv_i) \geq 0$ for $i \in I\setminus\{i_0\}$. Hence $x^* \in N_{I\setminus\{i_0\}} \cap \text{ri}(N(x|C))$, a contradiction to (B.1). Therefore, $I_- = \emptyset$ and similarly $\{i : v_i > 0\} = \emptyset$. Consequently, the vectors $\{\bar{c}_i : i \in I\}$ are linearly independent yielding the linear independence of the vectors $\{c_i : i \in I\}$.

It remains to show that (c) holds. Suppose that there is an $i_0 \in \{1, \cdots, s\}$ such that $\overline{c}_{i_0} \neq 0$ and $i_0 \notin U_{j=1}^{\ell} E_j$. If $\overline{c}_{i_0} \in \mathrm{ri}\,(N(x|C))$, then $\{i_0\}$ identifies $F(x|C)$ by Part (1) of Theorem 5.1. In this case,

$$N(x|C) = \mathrm{cone}\,\{\overline{c}_{i_0}\} + S.$$

But then cone $\{\overline{c}_{i_0}\} = S^{\perp} \cap N(x|C)$ so that $\overline{c}_{i_0}$ is an extreme direction for $S^{\perp} \cap N(x|C)$, a contradiction. Hence $\overline{c}_{i_0} \notin \mathrm{ri}\,(N(x|C))$. Therefore, there exists a subset J of I containing at least two elements with $J \neq I$ such that

$$\overline{c}_{i_0} = \sum_J u_i \overline{c}_i$$

where $u_i > 0$ for $i \in J$. But then the set $\widehat{J} := [I \backslash J] \cup \{i_0\}$ identifies $F(x|C)$, by Part (1) of Theorem 5.1, since

$$x^* = \sum_{I \backslash J} \overline{c}_i + \overline{c}_{i_0} \in \mathrm{ri}(N(x|C)).$$

Consequently, $\widehat{J}$ strongly identifies $F(x|C)$. But $N(x|C) \neq N_{\widehat{J}} + S$ since $\overline{c}_i \notin N_{\widehat{J}}$ for $i \in J$. Hence no such $\overline{c}_{i_0}$ exists.

($\Longleftarrow$) If $\ell = 0$, then the result follows trivially.

Thus assume that $\ell \geq 1$. Suppose I identifies $F(x|C)$. Let $J := \{j \in \{1, \cdots, \ell\} : d_j = \lambda \overline{c}_i$ for some $i \in I\}$ and let $K := \mathrm{cone}\,\{d_j : j \in J\}$. Then

$$N_I + S = K + S. \tag{B.2}$$

By Part (1) of Theorem 5.1, $N_I \cap \mathrm{ri}\,(N(x|C)) \neq \emptyset$, so $K \cap \mathrm{ri}\,(N(x|C)) \neq \emptyset$ by (B.2).

Let $x^* \in K \cap \mathrm{ri}\,(N(x|C))$ and choose $\widehat{u}_j \geq 0$ $j \in J$ so that

$$x^* = \sum_J \widehat{u}_j d_j.$$

Also, $x^* \in \mathrm{ri}\,(\mathrm{cone}\,\{d_j : j = 1, \cdots, \ell\})$, since

$$x^* \in \mathrm{ri}(N(x|C)) \cap S^{\perp} = \mathrm{ri}(\mathrm{cone}\,\{d_j : j = 1, \cdots, \ell\}).$$

Hence there are scalars $u_j > 0, j = 1, \cdots, \ell$ such that $x^* = \sum_{j=1}^{\ell} u_j d_j$. But then

$$0 = \sum_J (u_j - \widehat{u}_j) d_j + \sum_{j \in J} u_j d_j.$$

The linear independence of the d_j's implies that $u_j = 0$ for $j \notin J$. The choice of the u_j's implies that $J = \{1, \cdots, \ell\}$. Consequently, I strongly identifies $F(x|C)$. □

The proof of Theorem 6.3 now follows immediately from the above lemma and the MFCQ by taking $C = L\Omega(x, 0)$.

Acknowledgments. The author gratefully acknowledges the detailed comments and suggestions of two referees. Their efforts have greatly contributed to the exposition in this article.

REFERENCES

[1] J.-P. AUBIN AND I. EKELAND, *Applied Nonlinear Analysis*, John Wiley and Sons, 1984.

[2] J. V. BURKE, *An exact penalization viewpoint of constrained optimization*, Argonne National Laboratory, Mathematics and Computer Science Division Report ANL/MCS-TM-95, Arronne, Illinois, 1987.

[3] ——— *A sequential quadratic programming method for potentially infeasible mathematical programs*, J. Math. Anal. Appl., 139 (1989), pp. 319–351.

[4] J. V. BURKE AND S.-P. HAN, *A robust sequential quadratic programming method*, Math. Programming, 43 (1989), pp. 277–303.

[5] J. V. BURKE AND J. J. MORÉ, *On the identification of active constraints*, SIAM J. Numer. Anal., 25 (1988), pp. 1197–1211.

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

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

[8] R. FLETCHER, *Practical Methods for Optimization*, Vol. II, *Constrained Optimization*, John Wiley, New York, 1981.

[9] ———, *A model algorithm for composite nondifferentiable optimization problems*, Math. Programming Study, 17 (1982), pp. 67–76.

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

[11] S.-P. HAN AND O. L. MANGASARIAN, *Exact penalty functions in nonlinear programming*, Math. Programming, 14 (1978), pp. 73–86.

[12] O. L. MANGASARIAN AND S. FROMOVITZ, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl., 17 (1967), pp. 37–47.

[13] M. J. D. POWELL, *A fast algorithm for nonlinearly constrained optimization calculations*, Proceedings of the 1977 Dundee Biennial Conference on Numerical Analysis, Springer-Verlag, Berlin, 1977.

[14] S. M. ROBINSON, *Stability theory for systems of inequalities, Part* II*: Differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[15] S. M. ROBINSON AND R. R. MEYER, *Lower semi-continuity of multivalued linearization mappings*, SIAM J. Control, 11 (1973), pp. 525–533.

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

[17] ———, *Lipschitzian properties of multifunctions*, Nonlinear Anal. Theory, Methods, Appl., 9 (1985), pp. 867–885.

[18] M. SAHBA, *Globally convergent algorithm for nonlinearly constrained optimization problems*, J. Optim. Theory Appl., 52 (1987), pp. 291–309.

[19] R. B. WILSON, *A simplicial algorithm for concave programming*, Ph.D. dissertation, Graduate School of Business Administration, Harvard University, Boston, MA, 1963.

[20] S. WRIGHT, *Convergence of SQP-like methods for constrained optimization*, SIAM J. Control Optim., 27 (1989), pp. 13–26.

[21] ———, *An inexact algorithm for composite nondifferentiable optimization*, preprint, Math. Dept., North Carolina State University, Raleigh, NC, 1987.

[22] 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, New York, 1971.