LOCAL STRUCTURE OF FEASIBLE SETS
IN NONLINEAR PROGRAMMING
PART II: NONDEGENERACY

Stephen M. Robinson

October 1983
CP-83-52

PREFACE

A very important element in the research project on Adaptation and Optimization is provided by visiting scientists invited to work with the group for periods of several months. This leads to a continual exchange of ideas and experiences, prevents stagnation, and contributes greatly to the development of IIASA's research in this field.

One of these visitors during the summer of 1983 was Prof. Stephen Robinson of the University of Wisconsin-Madison. While at IIASA he completed this paper, in which he examines the local structure of the feasible set of a nonlinear programming problem under the condition of nondegeneracy.

ANDRZEJ WIERZBICKI
Chairman
System and Decision Sciences

LOCAL STRUCTURE OF FEASIBLE SETS

IN NONLINEAR PROGRAMMING. PART II: NONDEGENERACY

Stephen M. Robinson

University of Wisconsin - Madison

ABSTRACT

In this paper we examine the local structure of the feasible set of a nonlinear programming problem under the condition of nondegeneracy. We introduce this condition, examine its relationships to known properties of optimization problems, and show that when it holds at a given point the portion of the feasible set near that point is diffeomorphic to a simple convex set (often polyhedral). Moreover, this diffeomorphic relation is stable under small changes in the problem functions.

0. Introduction.

In this paper we study the system of nonlinear constraints

$$h(x) = 0$$
$$x \in C \cap \Omega ,$$

(0.1)

where h is a C^r function from an open subset Ω of $\mathbb{R}^n$ to $\mathbb{R}^m$ ($r \geq 1$), and C is a nonempty closed convex set in $\mathbb{R}^n$. We shall be concerned with properties of the set F of points satisfying (0.1) under a certain regularity condition that we term nondegeneracy. This condition is defined in Section 1, where its relationships to standard constraint qualifications used in nonlinear programming are also explored.

In Section 2 we introduce perturbations into the function h, and we show that under nondegeneracy the portions of the feasible sets of (0.1), and of all problems "close" to (0.1), that are near x_o are mutually C^r diffeomorphic. Geometrically, this means that if we do not perturb h very much then we are essentially working with the same problem, up to a C^r transformation of the coordinates. Moreover, this common feasible set has, with the correct choice of coordinates, a very simple structure: it is just the intersection of the part of C near x_o with the tangent space to $h^{-1}(0)$ through x_o. The coordinates through which this simplification is made are easy to compute, and this means that nonlinear optimization problems whose feasible sets look like (0.1) can actually be regarded, locally, as problems of optimization on a fixed convex

set which will be polyhedral if C is polyhedral.

The main point of Section 3 is to apply this point of view to establish some properties of local minimizers of such problems. We show that all such local minimizers must lie on a certain face of C, so that the optimization problem in question may be further simplified by restricting one's attention to that face of C. Finally, we show that when the problem is thus restricted the local minimizer in question inherits from the original problem the property of nondegeneracy, so that all of the results established under the nondegeneracy hypothesis will remain valid for the restricted problem.

1. Nondegeneracy.

In [4] we introduced a fundamental decomposition of $\mathbb{R}^n$ around a feasible point for (0.1), and we used it to prove various results about the feasible set $F = C \cap h^{-1}(0)$. Since we shall use the same decomposition here, we first summarize its properties.

Given a point $x_0 \in F$, denote $h'(x_0)$ by D. Let M be the subspace of $\mathbb{R}^n$ parallel to the affine hull A of the set C, and assume $D(M) = \mathbb{R}^m$. Denote $M \cap \ker D$ by K, and let L and J be subspaces complementary to K in M and in $\ker D$ respectively. Then one has $\mathbb{R}^n = J \oplus K \oplus L$, and one can define the projectors P_J, P_K, and P_L onto the spaces indicated along, in each case, the other two spaces. One has $P_J + P_K + P_L = I$, and the product of any two of these three projectors is zero. We shall denote by P_0 the projector $P_J + P_K$ onto $\ker D$ along L.

With these definitions the restriction of D to L is a bijection onto $\mathbb{R}^m$, and this bijection has an inverse D^-: a linear operator from $\mathbb{R}^m$ to $\mathbb{R}^n$ whose image is L and for which $DD^- = I$ and $D^-D = P_L$. One way to compute D^- is to let E be any bijection from $\mathbb{R}^m$ to L, then set $D^- = E(DE)^{-1}$; the inverse of DE exists because we assumed $D(M) = \mathbb{R}^m$.

The condition of nondegeneracy that we introduce here deals with the relationship of the spaces just described to the convex set C appearing in (0.1). Geometrically, it says that the space L can be chosen so that it is complementary to the tangent space (ker D) to $h^{-1}(0)$ at x_0, while at the same time for any x in C near x_0, movement along L near x does not cause one to leave C.

DEFINITION 1.1:

Let x_0 satisfy (0.1). We say x_0 is a <u>nondegenerate feasible point</u> for (0.1) if $D(M) = \mathbb{R}^m$ and there exist an open neighborhood N of the origin in $\mathbb{R}^n$ and a subspace L, complementary to K in M, such that

$$C \cap (x_0 + N) + L \cap N \subset C . \tag{1.1}$$

It might appear that instead of (1.1) one could impose the simpler condition that $x_0 + L \cap N \subset C$. However, this condition is not equivalent to (1.1), and in fact we shall need (1.1). To see that the two conditions are not equivalent, consider the closed convex cone C in $\mathbb{R}^3$ defined by

$$C := \{(x_1, x_2, x_3) \mid x_1 \geq 0, \ x_3 \geq 0, \ x_2^2 \leq 2x_1x_3\} .$$

Evidently aff $C = \mathbb{R}^3$. If we let $x_0 := (0,0,1)$ and $K := \mathbb{R}^2 \times \{0\}$, then

we can let $L := \{0\}^2 \times \mathbb{R}$, and for an appropriate neighborhood N of

the origin in $\mathbb{R}^3$ we have $x_0 + L \cap N \subset C$. However, for any positive

ε we have $(\varepsilon^2, \varepsilon, 1) \in C$, but for all small positive δ, $(\varepsilon^2, \varepsilon, 1-\delta) \notin C$.

Hence the condition (1.1) is not satisfied.

It is not difficult to show that if L is complementary in M to

$M \cap \ker h'(x_0)$, then for x near x_0, L is also complementary to

$M \cap \ker h'(x)$. It follows easily that if x_0 is a nondegenerate feasible

point for (0.1) then so is each feasible point sufficiently close to x.

If x_0 is a nondegenerate feasible point for (0.1) then it is

necessarily also a regular point in the sense of [4]. To see this,

note that $L \oplus K = M$, so that $D(L) = D(M) = \mathbb{R}^m$. Therefore, for the

neighborhood N of (1.1) it must be true that $D(L \cap N)$ is a neighborhood

of the origin in $\mathbb{R}^m$. We can rewrite (1.1) as

$$(C - x_0) \cap N + L \cap N \subset C - x_0 \, ,$$

and by applying D to both sides we see that $D(C - x_0)$ is a neighborhood

of 0 in $\mathbb{R}^m$. But this is one of the equivalent ways of expressing

regularity [4, Prop. 1.1].

There are several ways in which (1.1) might be satisfied. One obvious

way is for C to be a cylinder of the form $Q + C'$, where Q is a subspace

and C' is convex. If $Q + K = M$, then the requirement is satisfied. Of

course, nondegeneracy may hold also under much less stringent conditions,

since (1.1) deals only with local behavior.

In case C is polyhedral, as it is in many practical applications, then the __tangent cone__ to C at x_0, $T_C(x_0)$, coincides with $C - x_0$ near the origin. In that case we can obtain a simple characterization of nondegeneracy in terms of $T_C(x_0)$ and its polar $N_C(x_0)$, the __normal cone__ to C at x_0. The statement of this characterization uses the idea of the __lineality space__ of a cone Z: this space, written $\lim Z$, is the largest subspace in Z (i.e., $Z \cap (-Z)$).

PROPOSITION 1.2:

__Suppose C is polyhedral and__ x_0 __is feasible for (0.1). Then each of the following is necessary and sufficient for__ x_0 __to be nondegenerate:__

   __a.__ $D[\lim T_C(x_0)] = \mathbb{R}^m$.

   __b.__ $\{u \mid D^*u \in \text{aff } N_C(x_0)\} = \{0\}$.

PROOF:

The two conditions are equivalent since

$$D[\lim T_C(x_0)] = D[\{\text{aff } N_C(x_0)\}^\perp] = \{u \mid D^*u \in \text{aff } N_C(x_0)\}^\perp .$$

To show that (a) implies nondegeneracy, we first observe that $\lim T_C(x_0) \subset M$, so under (a) we have $D(M) = \mathbb{R}^m$. Next, we note that to establish (1.1) it is enough to show that

$$\lim T_C(x_0) + K = M . \tag{1.2}$$

Indeed, if (1.2) holds we can take L to be any subspace of $\lim T_C(x_0)$ that is complementary to K in M. Then since C is polyhedral, $T_C(x_0)$ coincides with $C - x_0$ near the origin, so that (1.1) will hold for an appropriate N.

To establish (1.2) we need only show that $\lim T_C(x_0) + K \supset M$, since the reverse inclusion is trivial. Let $m \in M$; since $D[\lim T_C(x_0)] = \mathbb{R}^m$ there is some $\ell \in \lim T_C(x_0)$ with $D\ell = Dm$. But then $m - \ell \in K$, so $m = \ell + (m - \ell) \in \lim T_C(x_0) + K$, which proves (1.2).

We next show that nondegeneracy implies (a). We are given that $D(M) = \mathbb{R}^m$ and that a subspace L exists, complementary to K in M, for which (1.1) holds for some neighborhood N. In particular, it follows from (1.1) that $L \cap N \subset C - x_0$. If we write cone A for the cone generated by a set A, we then have

$$L = \text{cone } L \cap N \subset \text{cone } (C - x_0) \subset \text{cl cone } (C - x_0) = T_C(x_0),$$

so $L \subset \lim T_C(x_0)$. But we observed earlier that if nondegeneracy holds then $D(L) = \mathbb{R}^m$, so

$$\mathbb{R}^m = D(L) \subset D[\lim T_C(x_0)] ,$$

and this completes the proof.

At this point, it may be helpful to review some special cases in order to put the idea of nondegeneracy into perspective. First, let us consider the linear programming problem in which $C = \mathbb{R}^n_+$ and $h(x) = Ax - b$. If x_0 is a basic feasible point corresponding to a partition of A into $[B\ N]$, where B is a basis for $\mathbb{R}^m$, then we have

$$T_C(x_0) = \underset{i=1}{\overset{n}{X}} Q_i ,$$

where

$$Q_i = \begin{cases} \mathbf{R} & \text{if } (x_0)_i > 0 \\ \mathbf{R}_+ & \text{otherwise .} \end{cases}$$

However, since x_0 is a basic feasible point the only possible indices i for which $(x_0)_i$ could be positive are $1,\ldots,m$, and therefore $\text{lin } T_C(x_0)$ is contained in the subspace of the first m coordinates. But for nondegeneracy as expressed in part (a) of Proposition 1.2 to hold, we must have $\dim \text{lin } T_C(x_0) = m$: that is, $\text{lin } T_C(x_0)$ must be exactly the subspace of the first m coordinates. However, this in turn implies that $(x_0)_1,\ldots,(x_0)_m$ must all be positive, so we see that in this case nondegeneracy in the sense used here reduces to the standard concept of nondegeneracy in linear programming.

For another example, consider a system of nonlinear constraints of the type often seen in the literature:

$$\begin{aligned} g_1(z) &\leq 0 \\ g_2(z) &= 0 , \end{aligned} \tag{1.3}$$

where g_1 and g_2 are differentiable functions from an open set $W \subset \mathbf{R}^k$ into $\mathbf{R}^p$ and $\mathbf{R}^q$ respectively. To rewrite (1.3) in the form (0.1), let $n = k + p$ and $m = p + q$; write $x = (z,s)$ with

$$h(x) = h(z,s) = \begin{bmatrix} g_1(z) + s \\ g_2(z) \end{bmatrix}$$

and take $C = \mathbf{R}^k \times \mathbf{R}^p_+$. If z_0 is feasible for (1.3), then with $s_0 = -g_1(z)$ the point $x_0 = (z_0,s_0)$ satisfies $h(x_0) = 0$, $x_0 \in C$. We have

$$D = h'(x_0) = \begin{bmatrix} g_1'(z_0) & I \\ g_2'(z_0) & 0 \end{bmatrix} \quad,$$

and

$$\lim T_C(x_0) = \mathbb{R}^k \times (X_{i=1}^p Q_i) \quad,$$

with

$$Q_i = \begin{cases} \{0\} & \text{if } i \in A \\ \mathbb{R} & \text{if } i \notin A \end{cases} \quad,$$

where A is the set $\{i \in \{1,\ldots,p\} \mid g_1(z_0)_i = 0\}$ of indices corresponding to active inequality constraints. Therefore D will carry $\lim T_C(x_0)$ onto $\mathbb{R}^{p+q}$ if and only if the set $\{g_1'(z_0)_i \mid i \in A\} \cup \{g_2'(z_0)_j \mid 1 \le j \le q\}$ has rank equal to the sum of q and the cardinality of A: that is, if and only if the gradients of the constraints active at z_0 are linearly independent. Thus, here again nondegeneracy corresponds to a well known property for the special case under consideration.

One might reasonably ask whether, in the case of a polyhedral set C, a representation like (1.3) could be used instead of (0.1). That is, could we simply represent the set C by a system of linear inequality and/or equality constraints, adjoin those constraints to the equation $h(x) = 0$, and thereby obtain an equivalent system of the form (1.3)? Of course, the answer is yes insofar as this procedure maintains the same feasible set, but it is no in the sense that in the new system a given point may be degenerate whereas it was nondegenerate in the representation (0.1).

For an example to illustrate this difficulty, consider the system

$$\langle a,x \rangle = 0$$

$$x \in C ,$$

<div align="right">(1.4)</div>

where a is a vector in $\mathbf{R}^4$ with $a \neq 0$ and C is the polyhedral convex set defined by

$$C = \{(x_1,x_2,x_3,x_4) | x_3 \geq \max\{|x_1|,|x_2|\}\} .$$

The origin is feasible for (1.4), and we have

$$\lim T_C(0) = \{0\}^3 \times \mathbf{R} \subset \mathbf{R}^4 .$$

Evidently the image of $\lim T_C(0)$ under the linear operator $\langle a, \cdot \rangle$ is $\mathbf{R}$, so by Theorem 1.2 the origin is a nondegenerate feasible point for (1.4). However, if we try to represent C by a system of linear inequalities we obtain the system

$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix} x \leq 0$$

<div align="right">(1.5)</div>

$$\langle a,x \rangle = 0 .$$

It is not hard to show that (1.5) is a minimal representation of the feasible set in the sense of Telgen [6] . Thus, any representation must have at least five linear constraints, and all five must pass through the origin since the solution set of (1.5) is a cone. But then their gradients (in $\mathbf{R}^4$) cannot be linearly independent, so that the origin is a degenerate feasible point.

It follows that the representation (0.1) is indeed more general, at least for purposes of representing nondegeneracy, than is (1.3): that is, there exist feasible sets which can be represented in the form (0.1), whose nondegenerate points become degenerate under any equivalent representation of the form (1.3). This may indicate that (0.1) is a better choice than (1.3) for use as a "canonical" representation for the feasible set of a nonlinear programming problem.

2. Local structure of the feasible set.

Having the idea of nondegeneracy as introduced in Section 1, we can now proceed to study the structure of the feasible set near a non-degenerate point x_0. In fact, we are interested not only in the feasible set of the given problem but also in the feasible sets of all problems close to it in an appropriate sense. To make this idea precise we introduce a parameter p lying in some open subset Π of a real Banach space, and a function h: $\Omega \times \Pi \to \mathbb{R}^m$. The set we shall study is the set of $x \in \Omega$ satisfying

$$h(x,p) = 0, \quad x \in C , \qquad (2.1)(p)$$

and we denote the set of all such x by $F(p)$. To link this notation to the analysis of Section 1 we specify a point $p_0 \in \Pi$ and identify the "unperturbed" function $h(x,p_0)$ with the $h(x)$ of Section 1. In all of what follows we assume that $h(\cdot,\cdot)$ is C^r ($r \geq 1$).

By applying the implicit - function theorem to the function

$$F(x,y,p) := D^-h(x,p) + (I - D^-D)[x - (x_0 + y)]$$

using the hypothesis that $D(M) = \mathbb{R}^m$, one can prove the following theorem.

THEOREM 2.1:

If $D(M) = \mathbb{R}^m$, there exist neighborhoods U of the origin in $\mathbb{R}^n$, V of p_0 in Π, and W of x_0 in $\mathbb{R}^n$, and a unique function x: U × V → W, such that:

   i) $x(0,p_0) = x_0$,

and for each $y \in U$ and $p \in V$,

   ii) $x(y,p) \in A$ and $h[x(y,p),p] = 0$;

   iii) $P_0[x(y,p) - x_0] = P_0 y$.

Further, x is C^r on U × V.

We omit the proof since it is almost identical to the proof given in [4] ; the only difference is in the presence of the parameter p, which causes no essential change in the argument; of course, one has to appeal to a version of the implicit - function theorem for Banach spaces (e.g., [1, Th. 1(4.XVII)]). In many applications the space of perturbations may be finite - dimensional, but we have made it a Banach space here in order to cover situations in which it may be, e.g., a function space.

The main use we shall make of Theorem 2.1 will be a proof that if x_0 is a nondegenerate feasible point of $(2.1)(p_0)$, then for each p near p_0 the portion of $F(p)$ near x_0 is C^r diffeomorphic to the portion of the fixed convex set $\Lambda := K \cap (C - x_0)$ that is near the origin. Thus, under nondegeneracy the local structure of the sets $F(p)$ near x_0 does not change. The next theorem formulates this result.

THEOREM 2.2:

Assume x_0 is a nondegenerate feasible point of $(2.1)(p_0)$. Then there exist open neighborhoods $U_* \subset U$, $V_* \subset V$, and $W_* \subset W$, such that if for each $p \in V_*$ one defines

$$\psi_p := x(\cdot,p) | U_* \cap \Lambda$$

and

$$\theta_p := P_0[(\cdot) - x_0] | W_* \cap F(p) ,$$

then ψ_p is a C^r diffeomorphism of $U_* \cap \Lambda$ onto $W_* \cap F(p)$, and θ_p is its inverse.

PROOF:

Recall that since x_0 is a nondegenerate feasible point there is a neighborhood N of the origin in $\mathbb{R}^n$ such that

$$(C - x_0) \cap N + L \cap N \subset C - x_0 . \tag{2.2}$$

Shrink the neighborhoods U, V, and W of Theorem 2.1 to open neighborhoods U_*, V_*, and W' such that if $y \in K \cap U_*$, $p \in V_*$ and $z \in W'$ then $y \in N$, and such that U_* is open and convex with

a) $x(y,p) \in W'$ and $P_L[x(y,p) - x_0] \in N$,

and

b) $z - x_0 \in N$ and $-P_L(z - x_0) \in N$.

Finally, let $W_* = W' \cap [x_0 + P_0^{-1}(U_*)]$.

Choose any $p \in V_*$; we first show that $\psi_p : U_* \cap \Lambda \to W_* \cap F(p)$ and that $\theta_p \circ \psi_p$ is the identity of $U_* \cap \Lambda$. Suppose $y \in U_* \cap \Lambda$. Then as $P_0[x(y,p) - x_0] = y$, we have $x(y,p) - x_0 = (P_0 + P_L)[x(y,p) - x_0] =$
$= y + P_L[x(y,p) - x_0] \in (C - x_0) \cap N + L \cap N \subset C - x_0$. However, we already know $h[x(y,p)] = 0$, and we know $x(y,p) \in W'$ and $P_0[x(y,p) - x_0] = y \in U_*$, so $x(y,p) \in W_*$. Thus $x(y,p) \in W_* \cap F(p)$, so $\psi_p : U_* \cap \Lambda \to W_* \cap F(p)$. If we apply θ_p we find that $(\theta_p \circ \psi_p)(y) = \theta_p[x(y,p)] = P_0[x(y,p) - x_0] = y$, so $\theta_p \circ \psi_p$ is the identity of $U_* \cap \Lambda$.

Now suppose that $z \in W_* \cap F(p)$. Then $\theta_p(z) = P_0[z - x_0] =$
$= (I - P_L)(z - x_0) \in (C - x_0) \cap N + L \cap N \subset C - x_0$. However, since $z \in C$ we have $z - x_0 \in M$, so $P_0(z - x_0) \in K$, and since $z \in W_*$ we know $P_0(z - x_0) \in U_*$. Thus $\theta_p : W_* \cap F(p) \to U_* \cap \Lambda$. Now if we set $y = \theta_p(z)$ we have $z \in M + x_0 = A$, $h(z,p) = 0$ (since $z \in F(p)$), and $P_0(z - x_0) = \theta_p(z) = y$. By the local uniqueness assertion of Theorem 2.1 we then have $z = x(y,p) = (\psi_p \circ \theta_p)(z)$, so that $\psi_p \circ \theta_p$ is the identity of $W_* \cap F(p)$. This completes the proof.

Evidently, if p_1 and p_2 belong to V_* then $\psi_{p_2} \circ \theta_{p_1}$ is a C^r diffeomorphism of $W_* \cap F(p_1)$ onto $W_* \cap J(p_2)$ whose inverse is $\psi_{p_1} \circ \theta_{p_2}$. This provides a significant contrast to the situation of [4], in which only regularity (as opposed to nondegeneracy)

held. For example, consider the regular system

$$(1 \quad -1 \quad p)x = 0$$
$$x \in \mathbb{R}^2_+ \times \mathbb{R} ,$$

where $p_0 = 0$. The feasible set $F(p_0)$ is the half-plane
$\{(\alpha,\alpha,\beta)| \alpha \geq 0, \beta \in \mathbb{R}\}$, of dimension 2 in $\mathbb{R}^3$. The origin is
a degenerate feasible point. If p is now changed from 0 to small
positive or negative values, then the new feasible set $F(p)$
acquires a vertex at the origin and therefore cannot be diffeomorphic
to $F(p_0)$.

We can obtain a more detailed picture of the structure of $F(p)$
near x_0 if we examine the faces of Λ that meet U_*. Every point of Λ
is in the relative interior of exactly one face of Λ [5, Th. 18.2] .
The intersection of such a relative interior with the open set U_*
is again a relatively open convex set, so for each $p \in V_*$ its
image under ψ_p is a C^r manifold whose dimension equals that of the
face in question. The collection of all such manifolds must form
a partition of $W_* \cap F(p)$, so we have the following theorem:

THEOREM 2.3:

Assume the notation and hypotheses of Theorem 2.2. For each $p \in V_*$,
the set $W_* \cap J(p)$ is a union of mutually disjoint C^r manifolds, each
being the image under ψ_p of $U_* \cap$ ri F for a distinct face F of Λ.
If C is polyhedral then for each p there are only finitely many such
manifolds.

PROOF:

Everything has been proved except for the last statement, which follows from the fact that a polyhedral convex set has only finitely many faces [5, Th. 19.1].

3. Application to nonlinear programming.

Here we apply the results of Sections 1 and 2 to study the local solutions of the nonlinear optimization problem

$$\min_{x} f(x,p)$$
$$\text{subject to } h(x,p) = 0, \quad x \in C, \tag{3.1}(p)$$

where f is a C^r function from $\Omega \times \Pi$ into $\mathbf{R}$, and where h and C are as previously defined. Note that if x_0 is a nondegenerate feasible point for $(3.1)(p_0)$, then Theorem 2.2 says that for $p \in V^*$ the sets

$$W^* \cap C \cap \{x \mid h(x,p_0) = 0\}, \quad U^* \cap (C - x_0) \cap K$$

are C^r diffeomorphic copies of each other. Hence, if we are interested only in the behavior of $(3.1)(p)$ near x_0, we might just as well study the problem

$$\min_{y} \phi(y,p)$$
$$\text{subject to } y \in \Lambda, \tag{3.2}(p)$$

where we have set

$$\phi(y,p) = f[x(y,p),p] = [f(\cdot,p) \circ \psi_p] (y) .$$

Note that the feasible set of (3.2)(p) is a convex set, independent of p, which will be polyhedral if C is polyhedral. For the rest of this section we make the blanket assumption that C is polyhedral and x_0 is a nondegenerate feasible point of (3.1)(p_0).

We shall exploit the relationship between (3.1)(p) and (3.2)(p) in several different ways in what follows. One first result is a simple proof of the well known first order optimality condition. It will also yield geometric insight that we can apply to find other results.

PROPOSITION 3.1:

If x_0 is a local minimizer of (3.1)(p_0), then

$$f_x(x_0,p_0)P_0 \in -N_C(x_0) . \qquad (3.3)$$

PROOF:

Since x_0 is a local minimizer of (3.1)(p_0) the origin is a local minimizer of (3.2)(p_0). Suppose $d \in T_C(x_0)$; since $L \subset \text{lin } T_C(x_0)$ and $P_0 d = d - P_L d$, we have $P_0 d \in T_C(x_0)$. However, since $d \in T_C(x_0) \subset M$, we have $P_J d = 0$, and thus $P_0 d = P_K d \in K$. Hence $P_0 d \in K \cap T_C(x_0)$, and for small $\lambda > 0$ we have $\lambda P_0 d \in K \cap (C - x_0)$ since $(C - x_0)$ agrees with $T_C(x_0)$ near the origin. But then $\lambda P_0 d$ is feasible for (3.2)(p_0) and therefore $\phi(\lambda P_0 d, p_0) \geq \phi(0,p_0)$ for small λ. Hence we must have $\phi_y(0,p_0)P_0 d \geq 0$, but by the chain rule $\phi_y(0,p_0) = f_x(x_0,p_0)P_0$, and

since $P_0^2 = P_0$ we have $f_x(x_0,p_0)P_0d \geqq 0$ for all $d \in T_C(x_0)$. But this means $f_x(x_0,p_0)P_0 \in -N_C(x_0)$ and this completes the proof.

One may think that (3.3) is not the usual first order optimality condition. To see that it really is, recall that

$$P_0 = I - h_x(x_0,p_0)^- h_x(x_0,p_0)$$

and define

$$\lambda(x_0,p_0) := -[h_x(x_0,p_0)^-]^* f_x(x_0,p_0) . \qquad (3.4)$$

Then (3.3) becomes

$$f_x(x_0,p_0) + h_x(x_0,p_0)^* \lambda(x_0,p_0) \in -N_C(x_0) , \qquad (3.5)$$

which is the familiar multiplier rule.

It is worth noting that under the nondegeneracy hypothesis the multipliers appearing in (3.5) are unique. To see this, suppose that λ_1 and λ_2 satisfy (3.5) when substituted for $\lambda(x_0,p_0)$ there. Then

$$h_x(x_0,p_0)^*(\lambda_1 - \lambda_2) \in \text{aff } N_C(x_0) ,$$

and by part (b) of Proposition 1.2 we then have $\lambda_1 = \lambda_2$.

Now denote $\phi_y(0,p_0)$ by g_0. The necessary optimality condition in Proposition 3.1 shows that, if x_0 is a local minimizer of (3.1)(p_0), then x_0 belongs to the set

$$\partial \psi_C^*(-g_0) = \{x \in C | <-g_0,x> = \sup_{c \in C} <-g_0,c>\} .$$

It is easy to show that this set is a face of C; we shall denote
it by F_0. It is also easy to show that $(F_0 - x_0) \cap K$ is a face of
Λ, which we shall denote by Φ_0. What may be less obvious is that if
we apply ψ_p to Φ_0, we recover, locally, exactly the intersection of
F_0 with $h^{-1}(0)$. This is the content of the next result.

PROPOSITION 3.2:

If x_0 <u>is a local minimizer of</u> (3.1)(p_0), <u>then for any</u> $p \in V_*$,

$$\psi_p[U_* \cap \Phi_0] = W^* \cap F_0 \cap F(p) .$$

PROOF:

Suppose $y \in U_* \cap \Phi_0$; let $x = \psi_p(y) = x(y,p)$. By Theorem 2.2,
$x \in W_* \cap F(p)$. By Theorem 2.1, $P_0(x - x_0) = P_0 y$. By the definition
of Φ_0, we have $y \in K \cap (F_0 - x_0)$; thus,

$$0 = \langle g_0, y \rangle = \phi_y(0, p_0) y = f_x(x_0, p_0) P_0 y =$$

$$= f_x(x_0, p_0) P_0(x - x_0) = \langle g_0, x - x_0 \rangle ,$$

so $x \in F_0$.

For the opposite inclusion, suppose that $x \in W^* \cap F_0 \cap F(p)$. Let
$y = \theta_p(x) = P_0(x - x_0)$; then we know that $x = \psi_p(y)$ and $y \in U_* \cap \Lambda$
by Theorem 2.2; thus in particular $y \in K$ and $x_0 + y \in C$. However,
since $P_0^2 = P_0$ we have

$$0 = \langle g_0, x - x_0 \rangle = f_x(x_0, p_0) P_0(x - x_0) =$$

$$= f_x(x_0, p_0) P_0[P_0(x - x_0)] = \langle g_0, y \rangle ,$$

and thus $x_0 + y \in F_0$, which completes the proof.

It will turn out that the face Φ_0 and, therefore, its image $F_0 \cap h^{-1}(0)$, contain all of the local minimizers of (3.2)(p) and (3.1)(p) respectively. In order to establish this fact, we show first that the local minimizers of (3.1)(p) and (3.2)(p) are in 1 - 1 correspondence for each p.

PROPOSITION 3.3:

For any $p \in V_*$ and any $y \in U_* \cap \Lambda$, y is a local minimizer of (3.2)(p) if and only if $\psi_p(y)$ is a local minimizer of (3.1)(p).

PROOF:

Obvious, since ψ_p and its inverse θ_p are diffeomorphisms and $\phi(\cdot,p) = f(\cdot,p) \circ \psi_p$.

The next result shows that the local minimizers of (3.2)(p) and of the more tightly constrained problem

$$\min \phi(y,p)$$

$$(3.6)(p)$$

$$\text{subject to } y \in \Phi_0 \, ,$$

are identical. From now on, we assume that $f_{xx}(x,p)$ is continuous at (x_0,p_0) and that $r \geq 2$.

THEOREM 3.4:

There exist neighborhoods U_1 of the origin in $\mathbb{R}^n$ and V_1 of p_0 in Π, such that for any $(y,p) \in U_1 \times V_1$, y is a local minimizer of (3.2)(p) if and only if y is a local minimizer of (3.6)(p).

PROOF:

We will require the following lemmas, proved in the appendix.

LEMMA 3.5:

Let C be a polyhedral convex set in $\mathbb{R}^n$. Let $z_0 \in \mathbb{R}^n$ and define $Z_0 := \partial\psi_C^*(z_0)$. Then there is a neighborhood U of z_0 such that if $z \in U$ then $\partial\psi_C^*(z) = \partial\psi_{Z_0}^*(z)$.

This lemma says, geometrically, that the face of C picked out by z is the same as the face of Z_0 picked out by z, as long as z is close to z_0. A related result was proved in [2]. The lemma obviously makes essential use of the property of polyhedrality.

LEMMA 3.6:

Let C, z_0 and Z_0 be as in Lemma 3.5. Assume $Z_0 \neq \phi$, and for $x \in \mathbb{R}^n$ let z(x) be the projection of x on Z_0. Then there is a constant $\gamma > 0$ such that for each $x \in C$,

$$\langle z_0, x - z(x) \rangle \leq -\gamma \| x - z(x) \| .$$

For the proof of Theorem 3.4, note first that if y is a local minimizer of (3.2)(p) that is close to 0, and if p is close to p_0, then $y \in \partial\psi_\Lambda^*[-\phi_y(y,p)]$. Using Lemma 3.5 we conclude that $y \in \partial\psi_{\Phi_0}^*[-\phi_y(y,p)]$ and, in particular, that $y \in \Phi_0$. But as y is a local minimizer of (3.2)(p), it is evidently also a local minimizer of the more tightly constrained problem (3.6)(p). For the converse, we can deduce from our

hypothesis about f_{xx} that there exist a ball $2\rho B$ of radius 2ρ about the origin in $\mathbb{R}^n$, a neighborhood V_1 of p_0 in Π, and a constant λ, such that if $y_1, y_2 \in 2\rho B$ and $p \in V_1$ then

$$\phi(y_2, p) = \phi(y_1, p) + \phi_y(y_1, p)(y_2 - y_1) + \alpha(y_1, y_2, p) ,$$

with $|\alpha(y_1, y_2, p)| \leq \frac{1}{2}\lambda \| y_1 - y_2 \|^2$. Further, if ρ and V_1 are taken to be small enough, then $\rho \leq \lambda^{-1}\gamma$ and if $y \in 2\rho B$ and $p \in V_1$ then $y \in U_*$, $p \in V_*$, and $\| \phi_y(y, p) - \phi_y(0, p_0) \| < \frac{1}{2}\gamma$, where γ comes from Lemma 3.6 with $z_0 := -\phi_y(0, p_0)$. Now let $U_1 := \rho B$. Let $(y_0, p) \in U_1 \times V_1$ and suppose y_0 is a local minimizer of $\phi(\cdot, p)$ on Φ, so that for some positive η, if $y \in \Phi$ with $\| y_0 - y \| < \eta$ then $\phi(y, p) \geq \phi(y_0, p)$. Choose any $y \in \Lambda$ with $\| y - y_0 \| < \min\{\eta, \rho\}$. Let y_Φ be the projection of y on Φ; then as $y_0 \in \Phi$ we have $\langle y - y_\Phi, y_0 - y_\Phi \rangle \leq 0$, so that $\| y_0 - y_\Phi \|^2 + \| y - y_\Phi \|^2 \leq$ $\leq \| y - y_0 \|^2$. Hence $\| y \| \leq \| y - y_0 \| + \| y_0 \| < 2\rho$ and similarly $\| y_\Phi \| < 2\rho$. Now we have

$$\phi(y, p) = \phi(y_\Phi, p) + \phi'(y_\Phi, p)(y - y_\Phi) + \alpha ,$$

where $|\alpha| \leq \frac{1}{2}\lambda \| y - y_\Phi \|^2$. Therefore

$$\phi(y, p) - \phi(y_0, p) = [\phi(y, p) - \phi(y_\Phi, p)] + [\phi(y_\Phi, p) - \phi(y_0, p)] \geq$$

$$\geq \phi(y, p) - \phi(y_\Phi, p) = \tag{3.7}$$

$$= \phi_y(y_\Phi, p)(y - y_\Phi) + \alpha ,$$

since $\phi(y_\Phi, p) \geq \phi(y_0, p)$ because $y_\Phi \in \Phi$ with $\| y_\Phi - y_0 \| \leq \| y - y_0 \| < \eta$. Applying Lemma 3.6 with $z_0 = -\phi_y(0, p_0)$, we find that

$$\phi_y(0, p_0)(y - y_\Phi) \geq \gamma \| y - y_\Phi \| .$$

But also, since $y_\phi \in 2\rho B$ and $p \in V_1$,

$$\| \phi_y(y_\phi, p) - \phi_y(0, p_0) \| < \tfrac{1}{2} \gamma .$$

Hence we conclude that

$$\phi_y(y_\phi, p)(y - y_\phi) \geq \tfrac{1}{2} \gamma \| y - y_\phi \| .$$

From (3.7) we then have

$$\phi(y, p) - \phi(y_0, p) \geq \phi_y(y_\phi, p)(y - y_\phi) + \alpha \geq$$
$$\geq \tfrac{1}{2} \gamma \| y - y_\phi \| - \tfrac{1}{2} \lambda \| y - y_\phi \|^2 \geq 0 ,$$

since $\| y - y_\phi \| \leq \rho \leq \lambda^{-1} \gamma$. It follows that y_0 is a local minimizer of (3.2)(p), which completes the proof of Theorem 3.4.

We have thus shown that, for the purpose of optimization, when considering (3.2)(p) one may as well restrict one's attention to the face ϕ_0: i.e., to the problem (3.6)(p). (This explains, incidentally, the prominent appearance of ϕ_0 in the second-order optimality conditions.) The immediate corollary to this is that in considering (3.1)(p) one may look only at the image of ϕ_0 under ψ_p; as we showed in Proposition 3.2, this will, locally, be the restriction of $F(p)$ to the face F_0 of C picked out by $f_x(x_0, p_0)P_0$. Thus we could consider, instead of (3.1)(p), the more tightly constrained problem

$$\begin{array}{l} \min_x f(x,p) \\[4pt] \text{subject to } h(x,p) = 0 \\[4pt] \qquad\qquad x \in F_0 . \end{array} \qquad (3.8)(p)$$

By Proposition 3.2 the local minima (near x_0) of (3.8)(p) will be the images under ψ_p of those of (3.6)(p). The latter, in turn, are by Theorem 3.4 identical to those of (3.2)(p), which finally are the images under θ_p of those of (3.1)(p) by Proposition 3.3. We conclude that, for any p near p_0, the local minimizer of (3.1)(p) and of (3.8)(p) are the same.

It is worth pointing out here that if F is any face of C that contains x_0, then the nondegeneracy hypothesis actually implies that x_0 is a nondegenerate feasible point for the system

$$h(x,p_0) = 0, \quad x \in F, \tag{3.9}$$

in which C has been replaced by F. To see this, let U be an open ball about the origin in $\mathbb{R}^n$, small enough so that the neighborhood N in the definition of nondegeneracy contains U. If $f - x_0 \in (F - x_0) \cap U$, then

$$(f - x_0) + L \cap U \subset (F - x_0) \cap U + L \cap U \subset C - x_0 \; .$$

However, $f + L \cap U$ is then a relatively open convex subset of C, which meets F. Thus, by [5, Th. 18.1] we have $f + L \cap U \subset F$. But this implies

$$(F - x_0) \cap U + L \cap U \subset F - x_0 \; ,$$

so that x_0 is a nondegenerate feasible point for (3.9). In particular, if in place of F we put the face F_0, we find that x_0, as a feasible point of (3.8)(p_0), inherits the nondegeneracy that it had as a feasible point of (3.1)(p_0).

Appendix.

The purpose of this appendix is to prove the two lemmas on polyhedral convexity required for the proof of Theorem 3.4. For convenience we repeat here the statements of the lemmas.

LEMMA 3.5:

Let C <u>be a polyhedral convex set in</u> $\mathbb{R}^n$. <u>Let</u> $z_0 \in \mathbb{R}^n$ <u>and define</u> $Z_0 := \partial\psi_C^*(z_0)$. <u>Then there is a neighborhood</u> U <u>of</u> z_0 <u>such that if</u> $z \in U$ <u>then</u> $\partial\psi_C^*(z) = \partial\psi_{Z_0}^*(z)$.

PROOF:

We first show that for z near z_0 we have $\partial\psi_C^*(z) \subset \partial\psi_{Z_0}^*(z)$. Indeed, if $y \in \partial\psi_C^*(z)$ then for each $c \in C$, $<z, c - y> \leq 0$. This holds in particular if $c \in Z_0$, so if $y \in Z_0$ then $y \in \partial\psi_{Z_0}^*(z)$. Thus we have to show that there is a neighborhood of z_0 such that for all z in that neighborhood we have $\partial\psi_C^*(z) \subset Z_0$. If this were not so, there would be a sequence $\{z_n\}$ converging to z_0, such that for each n there exists $y_n \in \partial\psi_C^*(z_n)\backslash Z_0$. Each set $\partial\psi_C^*(z_n)$ is a face of C, but the set of such faces is finite since C is polyhedral [5, Th. 19.1] . Therefore some face G of C recurs infinitely often among the $\partial\psi_C^*(z_n)$. Let $g \in G$; then for infinitely many n the pair (z_n, g) belongs to the (closed) graph of $\partial\psi_C^*$. But then (z_0, g) belongs to this graph too, so $g \in \partial\psi_C^*(z_0)$ and thus in fact $G \subset \partial\psi_C^*(z_0) \subset Z_0$. However, this contradicts the existence of the y_n, so we see that $\partial\psi_C^*(z) \subset \partial\psi_{Z_0}^*(z)$ for all z in some neighborhood, say U_1, of z_0.

To prove $\partial\psi^*_{Z_0}(z) \subset \partial\psi^*_C(z)$ for z near z_0, we first obtain a formula for

the inverse of $\partial\psi^*_{Z_0}$: i.e., for N_{Z_0}. If $z \in Z_0$ then, near 0, $C - z$ and

$Z_0 - z$ agree with $T_C(z)$ and $T_{Z_0}(z)$ respectively. Also, if we denote by

L the halfline $\{-\lambda_{Z_0} \mid \lambda \geq 0\}$, then we know that

$$Z_0 - z = (C - z) \cap L^O \ .$$

Thus for some neighborhood V of the origin,

$$V \cap T_{Z_0}(z) = V \cap (Z_0 - z) = V \cap (C - z) \cap L^O =$$

$$= V \cap T_C(z) \cap L^O \ .$$

But since $T_{Z_0}(z)$ and $T_C(z) \cap L^O$ are cones, we then have

$T_{Z_0}(z) = T_C(z) \cap L^O$ and hence, by polyhedrality, $N_{Z_0}(z) = N_C(z) + L$.

Thus, for each $z \in C$ we have

$$N_{Z_0}(z) = \begin{cases} N_C(z) + L, & \text{if } z \in Z_0 \\ \emptyset, & \text{if } z \notin Z_0 \end{cases} \tag{A.1}$$

Having this representation for N_{Z_0}, we next show that, for small

elements of $N_{Z_0}(z)$, only small elements of L need be used in the

representation (A.1). Recall that N_C is constant on the relative

interior of any convex subset of C, and in particular this holds

for faces of C. But the relative interiors of the faces partition

C [5, Th. 18.2], and thus N_C takes only finitely many distinct values.

Each of these is a cone in $\mathbb{R}^n$: call them $K_1,\ldots,K_I$. For each i let

G_i be the convex polyhedral multifunction taking $t \in \mathbb{R}_+$ to $K_i - tz_0$.

Then $K_i + L = G_i(\mathbb{R}_+)$, so by [3, Prop. 2] there is some $\alpha_i > 0$ with

$$B \cap (K_i + L) = B \cap G_i(\mathbb{R}_+) = G_i([0, \alpha_i]) \ .$$

It follows that

$$(\alpha_i^{-1} B) \cap (K_i + L) \subset G_i([0,1]) \ ,$$

and if we let $\alpha := \max_{i=1}^{I} \alpha_i$, then $\alpha > 0$ and

$$(\alpha^{-1} B) \cap (K_i + L) \subset G_i([0,1]), \quad i = 1, \ldots, I \ . \tag{A.2}$$

Hence in order to represent elements of $K_i + L$ having norm not greater than α^{-1}, we do not need to use elements of L having length more than 1.

Now define a neighborhood U of z_0 by letting $U = U_1 \cap (Z_0 + \alpha^{-1} B)$. Let $z \in U$; then either $\partial \psi_{Z_0}^*(z)$ is empty or it contains some point, say y. In the latter case $z \in N_{Z_0}(y)$, so if $w \in Z_0$ then $\langle z, w - y \rangle \leq 0$. But then also $\langle z - z_0, w - y \rangle \leq 0$ since $\langle z_0, w - y \rangle = 0$ because w and y belong to Z_0. Thus $z - z_0 \in N_{Z_0}(y)$, so by (A.1) and (A.2) we have for some i, $z - z_0 \in G_i[0,1]$. Thus there is some $t \in [0,1]$ with

$$z - z_0 \in N_C(y) - t z_0 \ . \tag{A.3}$$

However, we also know that $(1-t) z_0 \in N_C(y)$, since $N_C(y)$ is a cone and $y \in \partial \psi_C^*(z_0)$. Combining this with (A.3) we find that $z \in N_C(y)$, so that $y \in \partial \psi_C^*(z)$. It follows that $\partial \psi_{Z_0}^*(z) \subset \partial \psi_C^*(z)$, and this completes the proof of Lemma 3.5.

LEMMA 3.6:

Let C, z_0 and Z_0 be as in Lemma 3.5. Assume $Z_0 \neq \phi$, and for $x \in \mathbb{R}^n$ let $z(x)$ be the projection of x on Z_0. Then there is a constant $\gamma > 0$ such that for each $x \in C$,

$$\langle z_0, x - z(x) \rangle \leqq -\gamma \| x - z(x) \| \quad .$$

PROOF:

Define an extended real valued function $Q: \mathbb{R}^n \to \mathbb{R}$ by:

$Q(x) := \langle -z_0, x - z(x) \rangle + \psi_C(x)$. Since the projector taking x to

$z(x)$ can be written as $(I + N_{Z_0})^{-1}$, it is a polyhedral multifunction

in the sense of [3] . However, Q can be constructed from this

projector through operations of addition and composition with

obviously polyhedral multifunction, so Q itself is polyhedral.

Note that the set of zeros of Q is precisely Z_0. Applying

[3, Corollary to Prop. 1] we see that for some positive γ and

δ, and all $x \in C$ with $d[x, Z_0] := \inf \{ \| x - w \| \, | w \in Z_0 \} < \delta$, one

has

$$d[x, Z_0] = d[x, Q^{-1}(0)] \leqq \gamma^{-1} d[0, Q(x)] = \gamma^{-1} |\langle -z_0, x - z(x) \rangle| \quad . \qquad (A.4)$$

Choose any $x \in C$. For small positive μ the point $x_\mu := (1 - \mu) z(x) + \mu x$

remains in C and satisfies $d[x_\mu, Z_0] < \delta$. By (A.4), we then have

$$d[x_\mu, Z_0] \leqq \gamma^{-1} |\langle -z_0, x_\mu - z(x) \rangle| = \mu \gamma^{-1} |\langle -z_0, x - z(x) \rangle| \quad . \qquad (A.5)$$

However, one has for each $w \in Z_0$,

$$\langle x_\mu - z(x), w - z(x) \rangle = \mu \langle x - z(x), w - z(x) \rangle \leqq 0 \quad ,$$

so $z(x)$ is the projection of x_μ on Z_0. Hence

$$d[x_\mu, Z_0] = \| x_\mu - z(x) \| = \mu \| x - z(x) \| \quad ,$$

and from (A.5) we then obtain

$$\| x - z(x) \| \leq \gamma^{-1} |<-z_0, x - z(x)>| \ . \tag{A.6}$$

However, we note that since $z(x) \in \partial \psi_C^*(z_0)$, we have $<-z_0, c - z(x)> \geq 0$ for each $c \in C$ and particularly for $c = x$. Thus (A.6) becomes

$$<z_0, x - z(x)> \leq -\gamma \| x - z(x) \| \ ,$$

which completes the proof.

REFERENCES

[1]   L.V. Kantorovich and G.P. Akilov, Functional Analysis in
      Normed Spaces. Macmillan, 1959.

[2]   S.M. Robinson, "Strongly regular generalized equations,"
      Math. Operations Res. 5 (1980) 43 - 62.

[3]   S.M. Robinson, "Some continuity properties of polyhedral
      multifunctions," Math. Programming Studies 14 (1981) 206 - 214.

[4]   S.M. Robinson, "Local structure of feasible sets in nonlinear
      programming, Part I: Regularity," Technical Summary Report
      No. 2459, Mathematics Research Center, University of Wisconsin -
      Madison, 1982; forthcoming in Proceedings of the Inter-American
      Workshop on Numerical Methods, eds. V. Pereyra et al.,
      Springer-Verlag 1983.

[5]   R.T. Rockafellar, Convex Analysis. Princeton University Press,
      Princeton, NJ, 1970.

[6]   J. Telgen, "Minimal representation of convex polyhedral sets,"
      J. Optimization Theory Appl. 38 (1982).