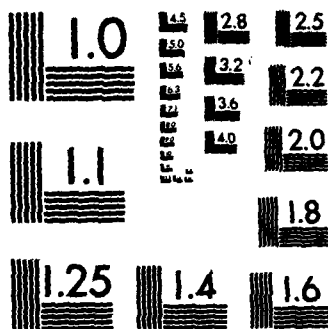MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

AD A125244

MRC Technical Summary Report #2459

LOCAL STRUCTURE OF FEASIBLE SETS IN
NONLINEAR PROGRAMMING, PART I:
REGULARITY

Stephen M. Robinson

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53706

December 1982

(Received November 24, 1982)

Approved for public release
Distribution unlimited

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

LOCAL STRUCTURE OF FEASIBLE SETS IN NONLINEAR PROGRAMMING,
PART I: REGULARITY*

Stephen M. Robinson**

Technical Summary Report #2459
December 1982

ABSTRACT

Given a feasible point for a nonlinear programming problem, we investigate the structure of the feasible set near that point. Under the constraint qualification called regularity, we show how to compute the tangent cone to the feasible set, and to produce feasible arcs with prescribed first and second derivatives. In order to carry out these constructions, we show that a particular way of representing the feasible set (as a system of equations with constrained variables) is particularly useful. We also give fairly short proofs of the first-order and second-order necessary optimality conditions in very general forms, using the arc constructions mentioned above.

SIGNIFICANCE AND EXPLANATION

Nonlinear programming is concerned with the problem of minimizing a function, often fairly smooth, over a set (the so-called feasible set) described by nonlinear inequality and equality constraints, as well as perhaps some bounds or other restrictions on the variables. Problems of nonlinear programming arise in statistics, in chemical engineering, in economics, and in many other areas.

Certain basic conditions, called optimality conditions, must be satisfied by any candidate for a solution of such a problem, provided that the constraints satisfy a reasonable regularity condition. These conditions describe the relationship of the derivatives of the function being minimized to the derivatives of the constraint functions and the set over which the minimization is being done. They form the basis for most numerical algorithms for solving such problems.

This paper examines the structure of the feasible set, and introduces some effective and fairly simple ways of dealing with this set. As a by-product of these techniques, comparatively simple and straightforward proofs of the optimality conditions and of related results are given.

LOCAL STRUCTURE OF FEASIBLE SETS IN NONLINEAR PROGRAMMING, PART I: REGULARITY[*]

Stephen M. Robinson[**]

1. <u>Introduction: the regularity condition</u>.

This paper deals with solution sets of systems

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

(1.1)

where h is a C^r function $(r \geq 1)$ from an open set Ω in $\mathbb{R}^n$ to $\mathbb{R}^m$, and where C is a convex set, not necessarily closed, in $\mathbb{R}^n$. We shall denote the set of all solutions of (1.1) by $C \cap h^{-1}(0)$, or briefly by F. One of the main application areas in which sets like F arise is nonlinear programming, since the constraints of many nonlinear programming problems either look like (1.1) or can be made to look that way by simple manipulations such as adding slack variables. Therefore, most of our analysis of (1.1) will be aimed at establishing results useful in nonlinear programming.

Given a point $x_0 \in F$, we often want to know what that part of F near x_0 is like. However, it is hard to tell much about F by looking directly at the nonlinear system (1.1). Therefore, a natural idea is to simplify (1.1) by linearizing h about x_0 and then to consider the system

$$h(x_0) + h'(x_0)(x - x_0) = 0$$
$$x \in C .$$

(1.2)

Since $h(x_0) = 0$ by assumption, the set of solutions of (1.2) is just
$x_0 + (C - x_0) \cap \ker h'(x_0)$, where $\ker h'(x_0)$ denotes the set $\{z | h'(x_0)z = 0\}$. We shall
denote this solution set by LF. Note that LF is a convex set, so we can deal with it
considerably more easily than we can with F.

An implicit assumption behind the construction and use of LF is that near x, LF is
"a good approximation" of F: one thinks, for example, of the tangent space to a smooth
manifold. However, it can easily happen that LF is nothing like F. Consider, for
example, the case with $n = 2$ and $m = 1$, in which

$$h(\xi, \eta) := \xi^2 - \eta$$

and

$$C := \{(\xi, \eta) | \eta \leq 0\} .$$

With these definitions, it is clear that F is the origin in $\mathbb{R}^2$, while if we take x_0
to be the origin then $LF = \mathbb{R} \times \{0\} = \{(\lambda, 0) | \lambda \in \mathbb{R}\}$. On the other hand, if we take the
same function h but change C to $\{(\xi, \eta) | \xi^2 - \eta \leq 0\}$, then F becomes
$\{(\xi, \eta) | \xi^2 - \eta = 0\}$ while LF becomes the origin. Hence in these two cases F and LF
are not at all alike, and in general we will need some criterion to ensure that LF is
locally similar to F if we are to use LF to draw conclusions about F.

The criterion we shall use is an extension of the familiar requirement, for systems of
equations, that the derivative of h at x_0 carry $\mathbb{R}^n$ onto $\mathbb{R}^m$. The extension consists
in taking appropriate account of the presence of the convex set C, and to do this we need
the idea of a tangent cone. Since we shall use this idea later too, we introduce it in a
fairly general form: if S is a subset of $\mathbb{R}^n$ and $x \in \text{cl } S$, then the (Bouligand)
tangent cone to S at x is the set $T_S(x)$ consisting of all right derivatives of arcs
emanating from x with the property that every neighborhood of x meets the intersection
of S with the arc. It is easy to show that a point d belongs to $T_S(x)$ if and only if
there are sequences $\{s_n\} \subset S$, converging to x, and $\{\lambda_n\} \subset (0, +\infty)$, such that
$\lambda_n(s_n - x)$ converges to d. See, e.g., [2] for more information about tangent cones.

Having the tangent cone, we can now make the extension we mentioned earlier, by
defining regular points for (1.1).

-2-

DEFINITION: Let $x_0 \in C$ with $h(x_0) = 0$. Then x_0 is a <u>regular point</u> of (1.1) if $h'(x_0)$ carries $T_C(x_0)$ onto $\mathbb{R}^m$.

It will be helpful to have some equivalent forms of regularity available. Three are given in the next proposition.

PROPOSITION 1.1: <u>Let</u> h <u>and</u> C <u>be as previously defined, with</u> $x_0 \in F$. <u>Then the following are equivalent</u>:

$$\text{(a)} \qquad h'(x_0)[T_C(x_0)] = \mathbb{R}^m \qquad\qquad\qquad (1.3)$$

$$\text{(b)} \qquad h'(x_0)(\text{aff } C) = \mathbb{R}^m \ \underline{\text{and}} \ \ker h'(x_0) \cap \text{ri}(C - x_0) \neq \emptyset , \qquad (1.4)$$

<u>where</u> ri <u>denotes relative interior (see</u> [4] <u>for definitions)</u>.

$$\text{(c)} \qquad 0 \in \text{int}\{h(x_0) + h'(x_0)(C - x_0)\} , \qquad\qquad (1.5)$$

<u>where</u> int <u>denotes interior</u>.

Part (c) of Proposition 1.1 is the form in which regularity was introduced in [3]; part (b) is a generalized version of the well known Mangasarian-Fromovitz constraint qualification [1], and part (a) is used in [2]. For our constructions in this paper, we shall depend primarily on (b).

PROOF: Denote $h'(x_0)$ by D; let B be the unit ball of $\mathbb{R}^n$ and denote its intersection with $\text{aff}(C - x_0)$ by B_C.

(a $\Longrightarrow$ b): We know $T_C(x_0) = \text{cl cone}(C - x_0)$ since C is convex; in particular $T_C(x_0) \subset \text{aff}(C - x_0)$ so that $D(\text{aff } C) = \mathbb{R}^m$. If $(\ker D) \cap \text{ri}(C - x_0)$ is empty, then the principal separation theorem [4, Th. 11.3] guarantees the existence of a nonzero $w \in \mathbb{R}^m$ such that for each $c \in C$, $\langle wD, c - x_0 \rangle \leq 0$. But then for any $z \in \text{cl cone}(C - x_0) = T_C(x_0)$ we have $\langle w, Dz \rangle = \langle wD, z \rangle \leq 0$, so that $D[T_C(x_0)] \neq \mathbb{R}^m$, contradicting (a). It follows that $(\ker D) \cap \text{ri}(C - x_0) \neq \emptyset$, which proves (b).

(b $\Longrightarrow$ c): If $D(\text{aff } C) = \mathbb{R}^m$ then we know D is an open mapping, so there is a bounded neighborhood N of the origin in $\text{aff}(C - x_0)$ such that DN is a neighborhood of the origin in $\mathbb{R}^m$. Let $y \in (\ker D) \cap \text{ri}(C - x_0)$; then for some positive ε, $x_0 + y + \varepsilon B_C \subset C$. Choose a positive δ small enough that $\delta N \subset \varepsilon B_C$; then $x_0 + y + \delta N \subset C$ and thus $y + \delta N \subset C - x_0$. But $D[y + \delta N] = \delta DN$, a neighborhood of the

-3-

origin, and therefore $D(C - x_0)$ is also a neighborhood of the origin. As $f(x_0) = 0$, this proves (c).

(c $\Longrightarrow$ a): As $T_C(x_0) \supset C - x_0$, (c) implies that $D[T_C(x_0)]$ is a neighborhood of the origin. But it is also a cone; hence it must be R^m. This completes the proof.

With the idea of a tangent cone and the criterion of regularity, we are in a position to investigate the relationships between LF and F. In Section 2 we introduce a special coordinate system that is particularly well suited for this investigation; then in Section 3 we apply this construction to show that F and LF have the same tangent cone at x_0, and to show further that feasible arcs can be constructed with prescribed tangents in LF. Finally, we apply these arcs to give simple proofs of necessary conditions for optimization on F.

2. Construction of a coordinate system; feasible arcs.

In this section we show how to set up a special coordinate system that is fitted to the structure of F near a regular point x_0. Denote aff C by A, and the subspace parallel to A by M: then we shall first use the implicit-function theorem to identify that part of $A \cap h^{-1}(0)$ near x_0 with that part of $M \cap \ker h'(x_0)$ near 0, under the regularity hypothesis. Then we shall show how this identification leads to a very simple way of constructing feasible arcs with prescribed derivatives. These arcs will be applied in the next section to prove results about tangent cones and optimality conditions.

First, we are going to decompose $\mathbb{R}^n$ in a way that employs the subspaces, M and $\ker h'(x_0)$, of special importance to us. Let us denote $h'(x_0)$ by D. Then since $D(M) = D(\text{aff } C) = \mathbb{R}^m$ by part (b) of Proposition 1.1, we must have

$$\{0\} = [D(M)]^\perp = \{u \mid D^* u \in M^\perp\} .$$

But then also $(\text{im } D^*) \cap M^\perp = \{0\}$, and taking orthogonal complements we have $(\ker D) + M = \mathbb{R}^n$. Denote $(\ker D) \cap M$ by K, and let J and L be subspaces complementary to K in $\ker D$ and M respectively: then we have

$$\mathbb{R}^n = J \oplus K \oplus L, \quad J \oplus K = \ker D, \quad K \oplus L = M . \qquad (2.1)$$

Given the decomposition (2.1), we denote by P_J, P_K and P_L the projectors from $\mathbb{R}^n$ onto the subspaces J, K, and L, in each case along the other two spaces. Thus $P_J + P_K + P_L = I$, and the product of any two of these projectors is the zero operator. We shall often write P_0 for $P_J + P_K$, the projector on $\ker D$ along L.

We now construct a particular generalized inverse of D that will aid us in applying the implicit-function theorem. Note first that from (2.1) we have $\mathbb{R}^n = (\ker D) \oplus L$, and that

$$D(L) = D(L \oplus K) = D(M) = \mathbb{R}^m .$$

It follows that D is a bijection from L onto $\mathbb{R}^m$, and the generalized inverse that we want is just the inverse of this bijection. This will be a linear operator D^- from $\mathbb{R}^m$ to $\mathbb{R}^n$ having the properties that DD^- is the identity of $\mathbb{R}^m$, and $D^- D$ is the projector P_L. To construct it, we can choose any bijection E from $\mathbb{R}^m$ to L and let $D^- := E(DE)^{-1}$. The existence of $(DE)^{-1}$ is guaranteed since L is independent of

ker D, and it is clear that DD^- is the identity of $\mathbb{R}^n$. To show that $D^-D = P_L$, note that D^-D annihilates ker D, while if $\ell \in L$, then $\ell = Ev$ for some $v \in \mathbb{R}^m$, and then

$$D^-D\ell = D^-DEv = E(DE)^{-1}(DE)v = Ev = \ell .$$

As $\mathbb{R}^n = (\ker D) \oplus L$ we must have $D^-D = P_L$, and it follows from this that $I - D^-D = P_0$.

The construction of these subspaces and projectors may be easier to understand if we relate the present situation to the well known case of linear programming. There, we have

$$h(x) = Ax - a, \quad C = \mathbb{R}^n_+, \quad h'(x_0) = A ,$$

where A is a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^M$ and $a \in \mathbb{R}^m$. The regularity hypothesis that we are using implies in particular that A has full row rank, and there is no loss of generality in assuming that A is partitioned as $[B \ N]$, with B $m \times m$ and nonsingular. Then the space of the first m components of $\mathbb{R}^n$ is clearly of dimension m and independent of ker A, and we shall take it to be L. Now for the bijection E we can choose $\begin{bmatrix} I_m \\ 0 \end{bmatrix}$, where I_m denotes the $m \times m$ identity matrix, and if we write D in place of A we have

$$D^- = E(DE)^{-1} = \begin{bmatrix} I_m \\ 0 \end{bmatrix} B^{-1} = \begin{bmatrix} B^{-1} \\ 0 \end{bmatrix} .$$

Hence

$$P_L = D^-D = \begin{bmatrix} B^{-1} \\ 0 \end{bmatrix} [B \ N] = \begin{bmatrix} I_m & B^{-1}N \\ 0 & 0 \end{bmatrix} ,$$

and

$$P_0 = I - P_L = \begin{bmatrix} 0 & -B^{-1}N \\ 0 & I_{n-m} \end{bmatrix} .$$

The reader will recognize the last $n - m$ columns of P_0 as being the edges of the feasible region along which the simplex method can move away from x_0.

Returning to the nonlinear situation, let us define a function $F : \Omega \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$F(x,y) := D^-h(x) + (I - D^-D)[x - (x_0 + y)] . \qquad (2.2)$$

As im $D^- = L$ and $I - D^-D = P_0$, the two major terms on the right-hand side of (2.2) lie in the independent spaces L and ker D. Accordingly, one has $F(x,y) = 0$ if and only if

(i) $D^-h(x) = 0$ (that is, $h(x) = 0$, since ker $D^- = \{0\}$), and

(ii) $P_0[x - (x_0 + y)] = 0$ (that is, $x - (x_0 + y) \in L$).

Of course, if $y \in K$ then (ii) implies that $P_0(x - x_0) = y$, so $x = x_0 + y + \ell$ for some $\ell \in L$. Thus, <u>for</u> $x \in \Omega$ <u>and</u> $y \in K$, $F(x,y) = 0$ <u>if and only if</u> $h(x) = 0$ <u>and</u> $x \in x_0 + y + L$.

If we differentiate (2.2) with respect to x, we find that $F_x(x_0,y) = I$ for any $y \in \mathbb{R}^n$. Thus, we may apply the implicit-function theorem to produce neighborhoods U of the origin in $\mathbb{R}^n$ and V of x_0, and a unique C^r function $x : U \to V$, such that $x(0) = x_0$ and, for each $y \in U$, $F(x(y),y) = 0$. This, in turn, meant that for each $m \in U \cap M$, $h[x(m)] = 0$ and $P_0[x(m) - x_0] = P_0m = P_Km$. To find the derivatives of this function x, we note that since $F(x(y),y) = 0$ we have

$$0 = \frac{d}{dy} F(x(y),y)\big|_{y=0} = F_x(x_0,0)x'(0) + F_y(x_0,0) = x'(0) - P_0,$$

where we have used the fact that $F_x(x_0,0) = I$. Therefore $x'(0) = P_0$. Now, if $r \geq 2$ we can differentiate again to obtain

$$0 = \frac{d^2}{dy^2} F(x(y),y)\big|_{y=0} = \frac{d}{dy} [F_x(x(y),y)x'(y) + F_y(x(y),y)]\big|_{y=0}$$

$$= F_{xx}(x_0,0)x'(0)x'(0) + F_{xy}(x_0,0)x'(0) + F_x(x_0,0)x''(0)$$

$$+ F_{yx}(x_0,0)x'(0) + F_{yy}(x_0,0) . \qquad (2.3)$$

However, F_{xy}, F_{yx}, and F_{yy} are all zero since the only appearance of y in (2.2) is in a linear term; also, we know $F_x(x_0,0) = I$, $F_{xx}(x_0,0) = D^-h''(x_0)$, and $x'(0) = P_0$. Therefore (2.3) yields for arbitrary s,t:

$$x''(0)(s)(t) = -D^-h''(x_0)(P_0s)(P_0t) . \qquad (2.4)$$

These derivatives can be further simplified if their arguments remain in ker D, since there P_0 acts like the identity.

The following theorem describes various properties of x, some of which we have noted informally above.

THEOREM 2.1: Suppose x_0 is a regular point of (1.1) and let D, J, K, L and M be as previously defined. Then there exist neighborhoods U of 0 in R^n and V of x_0 in R^n, and a unique C^r function $x : U \to V$ such that $x(0) = x_0$ and, for each $y \in U$, $h[x(y)] = 0$ and $P_0[x(y) - (x_0 + y)] = 0$. One has $x'(0) = P_0$ and, if $r \geq 2$, $x''(0)(s)(t) = -D^-h''(x_0)(P_0 s)(P_0 t)$ for $s, t \in R^n$.

Note that Theorem 2.1 yields $P_0[x(y) - x_0] = P_0 y$, so that if $y \in K$ then $P_0[x(y) - x_0] = y$, so that $x(y) \in x_0 + P_0[x(y) - x_0] + P_L[x(y) - x_0] \in x_0 + y + L \subset x_0 + M = \text{aff } C$. Thus x maps portions of K near 0 to portions of $(\text{aff } C) \cap h^{-1}(0)$ near x_0. However, we are really interested in $F = C \cap h^{-1}(0)$, so we shall next examine how to keep $x(y)$ in C instead of just in aff C. It turns out that a very easy way to do this is to investigate arcs in F. Our initial result involves only first derivatives.

THEOREM 2.2: Let h be C^1 and let x_0 be a regular point for (1.1). Let $d \in \text{ri } T_C(x_0)$. In order that there exist a C^1 arc $w(t)$ in $F = C \cap h^{-1}(0)$ with $w(0) = x_0$ and $\dot{w}(0) = d$, it is necessary and sufficient that $d \in \ker h'(x_0)$.

PROOF (necessity): For all small t we must have $h[w(t)] = 0$; thus
$$0 = \frac{d}{dt} h[w(t)]\big|_{t=0} = h'(x_0)\dot{w}(0) = h'(x_0)d .$$

(sufficiency): If $d \in \ker h'(x_0)$, then since $d \in \text{ri } T_C(x_0) \subset M$, we have $d \in K$. Also, since
$$d \in \text{ri } T_C(x_0) = \text{ri cl cone}(C - x_0) = \text{ri cone}(C - x_0) = \text{cone ri}(C - x_0) ,$$
there exist $\mu > 0$ and $\epsilon > 0$ with $\mu d \in \text{ri}(C - x_0)$ and $\mu d + \epsilon B_M \subset C - x_0$, where $B_M = B \cap M$, the unit ball in M, and where cone S denotes the cone generated by the set S. Now let $y(t) := td$; then with $w := x \circ y$ we have, for small non-negative t,
$$w(t) = x_0 + \dot{w}(0)t + o(t) . \tag{2.5}$$
But
$$\dot{w}(0) = x'(0)\dot{y}(0) = P_0 d = d .$$
Also, since $y(t)$ remains in K we have $h[w(t)] = 0$ and $w(t) \in \text{aff } C$ by Theorem 2.1; hence $w(t) - x_0 \in M$ and, since $d \in K (\subset M)$, also the function denoted by $o(t)$ in

-8-

(2.5) remains in M. Thus

$$w(t) - x_0 = [1 - \mu^{-1}t]0 + \mu^{-1}t[\mu d + \mu t^{-1}o(t)] \ ,$$

and for small non-negative t this is a convex combination of points of $C - x_0$, hence
itself a point of $C - x_0$ by convexity. Thus, for all small non-negative t, $w(t) \in C$,
and this completes the proof.

If we are willing to assume that h is C^2 instead of only C^1, we can obtain a
C^2 arc, and we can prescribe not only its first derivative but also its second.

THEOREM 2.3: <u>Let</u> h <u>be</u> C^2 <u>and let</u> x_0 <u>be a regular point for (1.1). Let</u>
$d \in C - x_0$ <u>and</u> $s \in ri \ T_C(x_0)$. <u>In order that there exist a</u> C^2 <u>arc</u> w(t) <u>in</u>
$F = C \cap h^{-1}(0)$ <u>with</u> $w(0) = x_0$, $\dot{w}(0) = d$, <u>and</u> $\ddot{w}(0) = s$, <u>it is necessary and sufficient</u>
<u>that</u> $d \in ker \ h'(x_0)$ <u>and</u>

$$h''(x_0)dd + h'(x_0)s = 0 \ . \tag{2.6}$$

PROOF (necessity): Theorem 2.2 tells us that we must have $d \in ker \ h'(x_0)$. For
(2.6), recall that $h[w(t)] = 0$ for small non-negative t, so

$$0 = \frac{d^2}{dt^2} h[w(t)]|_{t=0} = h''(x_0)dd + h'(x_0)s \ ,$$

where we have used the facts that $\dot{w}(0) = d$ and $\ddot{w}(0) = s$.

(sufficiency): Suppose $d \in ker \ h'(x_0)$ and (2.6) holds. For small non-negative t
let

$$y(t) := td + \frac{1}{2} t^2 P_0 s \ .$$

Note that $s \in ri \ T_C(x_0) \subset M = K \oplus L$, so $P_0 s \in K$. But
$d \in (C - x_0) \cap ker \ h'(x_0) \subset (K \oplus L) \cap (J \oplus K) = K$, so in fact $y(t) \in K$. Let
$w := x \circ y$. For small non-negative t we have $h[w(t)] = 0$ and $w(t) \in aff \ C$ by Theorem
2.1. Also,

$$\dot{w}(0) = x'(0)\dot{y}(0) = P_0 d = d \quad (since \ d \in K) \ ,$$

and

-9-

$$\ddot{w}(0) = x''(0)\dot{y}(0)\dot{y}(0) + x'(0)\ddot{y}(0)$$

$$= -D^- h''(x_0)dd + P_0^2 s$$

$$= -D^- [-h'(x_0)s] + P_0 s$$

$$= (P_L + P_0)s = s \ ,$$

where we have used (2.6), $P_0^2 = P_0$, $D^- D = P_L$, and $P_L + P_0 = I$. Hence w has the prescribed derivatives.

Now from Taylor's theorem we have for small non-negative t,

$$w(t) = x_0 + td + \frac{1}{2} t^2 s + r(t) \ , \tag{2.7}$$

where $r(t) = o(t^2)$. Note that since $w(t) \in$ aff C and since $d \in K(\subset M)$ and $s \in$ ri $T_C(x_0) \subset M$, we must have $r(t) \in M$ also. Since

$$s \in \text{ri } T_C(x_0) = \text{ri cone}(C - x_0) = \text{cone ri}(C - x_0) \ ,$$

there exist $\mu > 0$ and $\varepsilon > 0$ with

$$\mu s + \varepsilon B_M \subset C - x_0 \ ,$$

where, as before, $B_M = B \cap M$. Rewriting (2.7) as

$$w(t) - x_0 = [1 - t - \frac{1}{2} \mu^{-1} t^2]0 + td + \frac{1}{2} \mu^{-1} t^2 [\mu s + 2\mu t^{-2} r(t)] \ , \tag{2.8}$$

we see that for small non-negative t the right-hand side of (2.8) is a convex combination of points of $C - x_0$, and therefore $w(t) \in C$. But we know $h[w(t)] = 0$, so $w(t)$ is feasible, as required. This proves Theorem 2.3.

In this section we have shown how to use regularity to gain substantial amounts of information about the structure of F near a regular point x_0. In the final section we show how to use this information to compute the tangent cone to F at x_0 and to establish necessary conditions for optimization on F.

3. Applications: the tangent cone to F; optimality conditions.

In this section we apply the construction of Section 2 to compute $T_F(x_0)$ when x_0 is a regular point, and to give simple proofs of the general first-order and second-order necessary optimality conditions of nonlinear programming. First we consider the tangent cone.

THEOREM 3.1: Let h be C^1, and let x_0 be a regular point for (1.1). Then $T_F(x_0) = T_C(x_0) \cap \ker h'(x_0)$.

PROOF: As $F \subset C$, we have $T_F(x_0) \subset T_C(x_0)$. If $d \in T_F(x_0)$ then there is a sequence $\{x_n\} \subset F$ with $x_n \to x_0$ and, for some sequence $\{\lambda_n\} \subset (0,+\infty)$, $\lambda_n(x_n - x_0) \to d$. For each n we have

$$0 = h(x_n) = h(x_0) + h'(x_0)(x_n - x_0) + o(\|x_n - x_0\|) ,$$

so

$$0 = h'(x_0)\lambda_n(x_n - x_0) + (\lambda_n\|x_n - x_0\|)\|x_n - x_0\|^{-1} o(\|x_n - x_0\|) .$$

Taking the limit, we find $d \in \ker h'(x_0)$, and thus $T_F(x_0) \subset T_C(x_0) \cap \ker h'(x_0)$.

For the opposite inclusion, observe that regularity implies

$$\phi \neq ri(C - x_0) \cap \ker h'(x_0) \subset \text{cone } ri(C - x_0) \cap \ker h'(x_0)$$

$$= ri \text{ cl cone}(C - x_0) \cap \ker h'(x_0) = ri \, T_C(x_0) \cap \ker h'(x_0) .$$

Let $d \in ri \, T_C(x_0) \cap \ker h'(x_0)$. Using Theorem 2.2, construct a C^1 arc $w(t)$ in F with $w(0) = x_0$ and $\dot{w}(0) = d$. For large n, define $x_n := w(n^{-1})$. We have $x_n \in F$, and $x_n \to x_0$. Also, with $\lambda_n := n$ we have

$$\lambda_n(x_n - x_0) = n[w(n^{-1}) - w(0)] = n[\dot{w}(0)n^{-1} + o(n^{-1})] = d + no(n^{-1}) .$$

Taking the limit we find that $\lambda_n(x_n - x_0) \to d$, so $d \in T_F(x_0)$. Hence

$$ri \, T_C(x_0) \cap \ker h'(x_0) \subset T_F(x_0) . \qquad (3.1)$$

The left side of (3.1) is the nonempty intersection of two relatively open sets, hence is the relative interior of the intersection of their closures. Accordingly, we have

$$ri[T_C(x_0) \cap \ker h'(x_0)] \subset T_F(x_0) ,$$

and since $T_F(x_0)$ is closed we have

$$T_C(x_0) \cap \ker h'(x_0) \subset T_F(x_0) .$$

This completes the proof.

Theorem 3.1 yields a simple proof of the first-order necessary optimality conditions under the hypothesis of regularity. To see how to construct such a proof, consider the problem of minimizing f on F, where f is a function from Ω to R.

THEOREM 3.2: Let f and h be C^1, and assume x_0 is a regular point of (1.1). If x_0 is a local minimizer for f on F, then there is some $\lambda \in R^m$ with

$$f'(x_0) + h'(x_0)^* \lambda \in -N_C(x_0) .$$

Here the asterisk denotes the adjoint operator, and $N_C(x_0)$ is the normal cone to C at x_0 defined by

$$N_C(x_0) = T_C(x_0)^* = \{y \in R^n | \forall z \in T_C(x_0), \langle y,z \rangle \leq 0\} .$$

PROOF: If $d \in T_F(x_0)$, then there are sequences $\{x_n\} \subset F$ and $\{\lambda_n\} \subset (0,+\infty)$ with $x_n \to x_0$ and $\lambda_n(x_n - x_0) \to d$. For all large n, local optimality implies that

$$0 \leq f(x_n) - f(x_0) = f'(x_0)(x_n - x_0) + o(\|x_n - x_0\|) ,$$

so

$$0 \leq f'(x_0)\lambda_n(x_n - x_0) + (\lambda_n\|x_n - x_0\|)\|x_n - x_0\|^{-1} o(\|x_n - x_0\|) .$$

Taking the limit we find that $f'(x_0)d \geq 0$, and this shows that $f'(x_0) \in -N_F(x_0)$. Also, by Theorem 3.1 we have $T_F(x_0) = T_C(x_0) \cap \ker h'(x_0)$, so

$$N_F(x_0) = [T_C(x_0) \cap \ker h'(x_0)]^* = cl[N_C(x_0) + im\, h'(x_0)^*] .$$

However, as noted in the proof of Theorem 3.1 the regularity condition implies $[ri\, T_C(x_0)] \cap \ker h'(x_0) \neq \phi$, so by [4, Cor. 23.8.1] we find that $N_C(x_0) + im\, h'(x_0)^*$ is closed (since the normal cone to a closed convex cone at the origin is the polar of that cone). Hence, for some $\lambda \in R^m$ we have

$$f'(x_0) + h'(x_0)^* \lambda \in -N_C(x_0) ,$$

which completes the proof.

We can derive a second-order necessary condition for the problem just considered, under the same regularity hypothesis, if we are willing to assume that f and h are C^2. This is done in the next theorem, whose proof is based on an unpublished proof of Weinberger [5] for a somewhat different problem.

THEOREM 3.3: Let f and h be C^2, and suppose that x_0 is a regular point of (1.1). If x_0 is a local minimizer for f on F, then for each

$d \in (C - x_0) \cap \ker h'(x_0)$ __with__ $f'(x_0)d = 0$, __there exists__ $\lambda \in \mathbb{R}^m$ __such that__

$$f'(x_0) + h'(x_0)^*\lambda \in -N_C(x_0)$$

__and__

$$f''(x_0)dd + \langle \lambda, h''(x_0)dd \rangle \geq 0 \ .$$

We note that under the hypotheses of this theorem, if $d \in (C - x_0) \cap \ker h'(x_0)$ then the first order condition implies $f'(x_0)d \geq 0$. Hence we could have substituted "$f'(x_0)d \leq 0$" in the statement of the thoerem without changing anything.

PROOF: We first show that for each $s \in \text{ri } T_C(x_0)$ the system

$$h'(x_0)d = 0, \quad h''(x_0)dd + h'(x_0)s = 0$$

$$f'(x_0)d = 0, \quad f''(x_0)dd + f'(x_0)s < 0 \qquad\qquad (3.2)$$

$$d \in C - x_0 \ ,$$

is inconsistent. Indeed, suppose that $s \in \text{ri } T_C(x_0)$ and that (3.2) is consistent. By Theorem 2.3 there is a C^2 arc $w(t)$ in F with $w(0) = x_0$, $\dot{w}(0) = d$, and $\ddot{w}(0) = s$. Define $\phi := f \circ w$. Then $\phi(0) = 0$, $\dot{\phi}(0) = f'(x_0)\dot{w}(0) = f'(x_0)d = 0$, and

$$\ddot{\phi}(0) = f''(x_0)\dot{w}(0)\dot{w}(0) + f'(x_0)\ddot{w}(0)$$

$$= f''(x_0)dd + f'(x_0)s < 0 \ .$$

Hence for small t we have $w(t) \in F$ and

$$f[w(t)] = \phi(t) = \phi(0) + \dot{\phi}(0)t + \frac{1}{2}\ddot{\phi}(0)t^2 + o(t^2)$$

$$= f(x_0) + t^2[\frac{1}{2}\ddot{\phi}(0) + t^{-2}o(t^2)] < f(x_0) \ .$$

Since this contradicts the assumption that x_0 was a local minimizer for f on F, (3.2) must be inconsistent.

The remainder of the proof consists of a separation argument designed to translate the inconsistency of (3.2) into a positive statement about the existence of λ. Choose $d \in (C - x_0) \cap \ker h'(x_0)$ with $f'(x_0)d = 0$, and define a linear transformation $G : \mathbb{R}^{n+1} \to \mathbb{R}^{m+1}$ by

$$G(s,\sigma) := \begin{bmatrix} h'(x_0)s + \sigma h''(x_0)dd \\ \\ f'(x_0)s + \sigma f''(x_0)dd \end{bmatrix} \ .$$

Recall that $\text{ri } G[T_C(x_0) \times \mathbb{R}_+] = G\{\text{ri}[T_C(x_0) \times \mathbb{R}_+]\} = G\{[\text{ri } T_C(x_0)] \times [\text{int } \mathbb{R}_+]\}$. For any

-13-

$s \in \text{ri } T_C(x_0)$ and any $\sigma > 0$, the inconsistency of (3.2) for the given d and s implies that

$$G(s,\sigma) \notin \{0\}^m \times \text{int } R_- = \text{ri}[\{0\}^m \times R_-] .$$

Hence,

$$\text{ri } G[T_C(x_0) \times R_+] \cap \text{ri}[\{0\}^m \times R_-] = \phi .$$

By [4, Th. 11.3] there exist $\lambda_0 \in R^m$, and $\mu, \nu \in R$ with $(\lambda_0, \nu) \neq 0$, such that for each $z \in T_C(x_0)$, $\gamma \in R_+$, and $\delta \in R_-$,

$$\langle \lambda_0, h'(x_0)z + \gamma h''(x_0)dd \rangle + \nu[f'(x_0)z + \gamma f''(x_0)dd] \geq \mu \geq \nu\delta . \quad (3.3)$$

The form of (3.3) ensures that $\mu \geq 0$ and $\nu \geq 0$. Also, by choosing $\gamma = 0$ in (3.3) we obtain

$$\langle \lambda_0, h'(x_0)z \rangle + \nu f'(x_0)z \geq 0 \quad (3.4)$$

for each $z \in T_C(x_0)$, and by choosing $z = 0$ and $\gamma = 1$ we have

$$\langle \lambda_0, h''(x_0)dd \rangle + \nu f''(x_0)dd \geq 0 . \quad (3.5)$$

If $\nu = 0$ then $\lambda_0 \neq 0$ and, from (3.4), for each $z \in T_C(x_0)$ we have $\langle \lambda_0, h'(x_0)z \rangle \geq 0$. But this means that $h'(x_0)[T_C(x_0)] \neq R^m$, which contradicts the regularity assumption (Proposition 1.1). Hence $\nu \neq 0$, so in fact $\nu > 0$. If we define $\lambda := \lambda_0/\nu$ then (3.4) yields $f'(x_0) + h'(x_0)^*\lambda \in -N_C(x_0)$, and (3.5) yields

$$f''(x_0)dd + \langle \lambda, h''(x_0)dd \rangle \geq 0 .$$

This completes the proof.

One might wonder whether we could replace $C - x_0$ in the statement of Theorem 3.3 by $T_C(x_0)$. In general, the answer is no. For example, consider $f : R^2 \rightarrow R$ given by $f(x_1, x_2) := x_2 - \frac{1}{2}x_1^2$, with $C \subset R^2$ defined by $C := \{(x_1, x_2)|x_2 \geq |x_1|^{3/2}\}$, and let h be vacuous. With $x_0 := (0,0)$ we have $T_C(x_0) = R \times R_+$. However, if we take $d = (1,0)$ then we have $f'(x_0)d = 0$ and $f''(x_0)dd = -1 < 0$, yet the origin is a local minimizer of f on C. The problem here is that there is no feasible C^2 arc emanating from the origin with tangent $(1,0)$. Of course, if C were polyhedral then, near x_0, $C - x_0$ would agree with $T_C(x_0)$, so in that case we could take $d \in T_C(x_0)$.

It is also not very difficult to find examples to show that one cannot in general use the same λ for every d in Theorem 3.3. However, if the λ appearing in the first

order condition should happen to be unique then Theorem 3.3 would guarantee that that λ would work for all d. This uniqueness is in fact realized in an important special case, that of nondegeneracy. That special case will be treated in detail in Part II of this paper.

REFERENCES

[1] 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.

[2] J.-P. Penot, On regularity conditions in mathematical programming. Math. Programming Studies 19 (1982), pp. 167-199.

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

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

[5] H. F. Weinberger, Conditions for a local Pareto optimum. Typescript, ca. 1974.

SMR:scr

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
2459	AD A125244	

4. TITLE (and Subtitle)	5. TYPE OF REPORT & PERIOD COVERED
LOCAL STRUCTURE OF FEASIBLE SETS IN NONLINEAR PROGRAMMING, PART I: REGULARITY	Summary Report - no specific reporting period
	6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)	8. CONTRACT OR GRANT NUMBER(s)
Stephen M. Robinson	MCS 7901066 (Mod. 2) DAAG29-80-C-0041 MCS 8200632

9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Mathematics Research Center, University of 610 Walnut Street Wisconsin Madison, Wisconsin 53706	Work Unit Number 5 - Mathematical Programming

11. CONTROLLING OFFICE NAME AND ADDRESS	12. REPORT DATE
See Item 18 below.	December 1982
	13. NUMBER OF PAGES
	16

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonlinear inequalities
nonlinear programming
nonconvex programming

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Given a feasible point for a nonlinear programming problem, we investigate the structure of the feasible set near that point. Under the constraint qualification called regularity, we show how to compute the tangent cone to the feasible set, and to produce feasible arcs with prescribed first and second derivatives. In order to carry out these constructions, we show that a particular way of representing the feasible set (as a system of equations with constrained variables) is particularly useful. We also give fairly short proofs of the first-order and second-order necessary optimality conditions in very general forms, using the arc constructions mentioned above.

END