MATHEMATICS OF OPERATIONS RESEARCH
Vol. 4, No. 4, November 1979
Printed in U.S.A.

THE GENERIC NATURE OF OPTIMALITY CONDITIONS IN NONLINEAR PROGRAMMING*

J. E. SPINGARN† AND R. T. ROCKAFELLAR‡#

First- and second-order optimality conditions are treated for parameterized classes of nonlinear programming problems in R^n. Under certain assumptions, it is shown that almost all problems in the class are such that every local minimizer satisfies the strong form of the optimality conditions.

1. The motivation behind generic conditions. Optimality conditions in nonlinear programming are used in the design and analysis of algorithms, the study of sensitivity to changes in parameter values, and in theoretical situations involving duality and the existence of "prices." It is characteristic of such applications that, to get anywhere, one has to assume that any solutions to the problem satisfy conditions stronger than are really *necessary* for optimality. However, such an assumption is awkward because it may well be impossible to verify.

For instance, an algorithm may be guaranteed to converge in a certain manner to a solution to a nonlinear programming problem, provided that the solution has certain properties. But these properties often cannot be checked (and the algorithm therefore justified) unless the exact solution is known in advance! In fact, it is rare in nonlinear programming that one ever sees more than an approximate solution, and then only after lengthy computation.

Actually, for most applications, the assertions that can be made about a *class* of problems are of greater value than those concerning a particular problem (in which numerical values are assigned to all parameters). One would like to know that a property or kind of behavior is "typical," although there may be "exceptional" cases. Intuition and experience are often helpful in this regard, but a sound theory requires a more solid mathematical footing. To draw an analogy, intuition and experience say a lot about how a computational procedure may be expected to work in practice, but theoretical analysis of convergence is valuable nonetheless for the insights and discipline it provides.

A condition is said to be *generic* for a class of problems if it is "usually" true for problems in the class in one of the several rigorous mathematical senses that have been developed. The aim of this paper is to treat the standard optimality conditions in nonlinear programming from this point of view. For simplicity, we limit ourselves here to the problem

$$\begin{aligned} &\text{minimize } f(x) \text{ over all } x \in R^n \text{ such that} \\ &g_i(x) \leqslant 0 \text{ for } i = 1, \ldots, m, \end{aligned} \tag{Q}$$

where the functions f and g_i are continuously differentiable.

Any theory of generic conditions depends heavily on the "classes" of problems relative to which the generic assertions are made, as well as the particular notion of

* Received April 6, 1978; revised October 26, 1978.
AMS 1970 subject classification. Primary 90C30.
IAOR 1973 subject classification. Main: Nonlinear programming.
OR/MS Index 1978 subject classification. Primary 653 Programming, nonlinear, theory, optimality conditions.
Key words: Nonlinear programming, optimality conditions, Sard's theorem.
† Department of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332.
‡ Department of Mathematics, University of Washington, Seattle, Washington 98195.
# Research partially supported by USAF grant number 77-3204.

0364-765X/79/0404/0425$01.25

"generic" which is adopted. One approach, reminiscent of work that has been done on various economic models, would be to identify (Q) with a point $(f, g_1, \ldots, g_m)$ in a linear space of highly differentiable functions from $\mathbf{R}^n$ to $\mathbf{R}^{m+1}$ under the Whitney topology. (In this topology, a sequence of functions converges if all the partial derivatives up to a certain order converge uniformly on all bounded sets.) A "class" of problems would be represented by such a space itself, or perhaps by some subset with nonempty interior. A condition would be said to hold *generically* if the subclass for which it fails forms a point set of first Baire category (i.e., a countable union of nowhere dense sets).

We feel this approach would be unsatisfactory in nonlinear programming for several reasons. The emphasis on high orders of differentiability seems alien to the subject. While smoothness may be necessary for certain results, it would seem to be a mistake to adopt a conceptual framework inherently dependent on it. Furthermore, the "classes" of problems are not sharp enough. (The deformations that are allowed are too general.) The associated notion of what constitutes an "exceptional" set of problems is very difficult to interpret in a meaningful way in applications.

The alternative that to us appears superior is an approach based on the "almost everywhere" concept of measure theory. In this case the "classes" of problems are much more restricted and better able to reflect the fine structure that must be taken into consideration. We imagine as a "class" a set of problems

$$\begin{aligned} &\text{minimize } f^w(x) \text{ over all } x \in R^n \text{ satisfying} \\ &g_i^w(x) \leqslant 0 \text{ for } i = 1, \ldots, m, \end{aligned} \tag{$Q(w)$}$$

where w is a parameter vector ranging over a set $W \subset R^k$. The components of w may represent costs, demand, etc., that are variable within the context of a particular model (as opposed to other parameter values that are "structural" and therefore best regarded as "fixed" for the class). It is natural to think of w as a *random* vector governed by some probability measure on W. An "exceptional" set of problems is then one whose probability is 0. A condition holds *generically* (relative to the given class) if it holds with probability 1.

We propose, therefore, the general study of *what assumptions are needed on a parameterized class of problems like* $\{(Q(w)) : w \in W\}$ *and a probability measure on* W *in order that various properties, such as standard conditions for optimality, are sure to hold except for a subclass of measure zero.* Of course this is a large project, and only a limited contribution can be presented here. Besides the motivation already mentioned, there is the prospect of specific applications to areas like stochastic programming, which are directly concerned with problems dependent on random parameters.

An important case is where the probability measure is given by a density with respect to ordinary k-dimensional (Lebesgue) measure on the set $W \subset R^k$; one can cover all such probabilities simultaneously by considering as *exceptional* those sets of Lebesgue measure zero. Spingarn's thesis [4] develops a fairly general theory addressed to this case and applicable to a wide choice of parameterizations.

In the present paper we treat only a special kind of parameterization by vectors $w = (v, u) \in R^n \times R^m$ which, however, turns out to be fundamental in some respects:

$$\begin{aligned} &\text{minimize } f(x) - x \cdot v \text{ over all } x \in R^n \\ &\text{satisfying } g_i(x) \leqslant u_i \text{ for all } i \in I = \{1, \ldots, m\}. \end{aligned} \tag{$Q(v, u)$}$$

Let us say that $(Q(v, u))$ is a "troublesome" problem if it has a locally optimal solution which does not satisfy the "strong second-order conditions for optimality" (cf. §3). Our main result states that when f and g are sufficiently smooth, the set of all

$w = (v, u) \in R^n \times R^m$ for which $(Q(v, u))$ is troublesome is of (Lebesgue) measure zero.

2. First-order conditions. For a feasible solution x to $(Q(v, u))$, the set of *active constraint indices* is

$$I(u, x) = \{ i \in I : g_i(x) = u_i \}.$$

The standard *first-order conditions* for local optimality of x in $(Q(v, u))$ are that x should be feasible and there should exist a vector $y \in R^m_+$ such that

$$\begin{aligned} &\nabla f(x) + \sum_I y_i \nabla g_i(x) = v \quad \text{and} \\ &\text{for all } i \in I, \quad \text{either } y_i = 0 \text{ or } i \in I(u, x). \end{aligned} \tag{KT}$$

Of course, as stated, these conditions are not actually *necessary* for optimality. They are only necessary under an additional assumption—a *constraint qualification*—that in practice is often difficult or impossible to verify. Many constraint qualifications have been studied, but the strongest that ever seems to be needed (for the purpose of drawing some conclusion about the first-order conditions) is

$$\{ \nabla g_i(x) : i \in I(u, x) \} \text{ is a linearly independent set.} \tag{CQ}$$

For the proof that the first-order optimality conditions are necessary when x satisfies (CQ), we refer to Hestenes [2].

THEOREM 1. *Suppose the functions $g_1, \ldots, g_m$ on R^n are of class C^r. Fix any $v \in R^n$. Then for all u except in a set of measure zero in R^m, $(Q(v, u))$ is such that every feasible solution x satisfies condition (CQ) (and hence every locally optimal solution x satisfies (KT) for some $y \in R^m_+$).*

The principal tool in the proof of theorem 1 will be the well-known theorem of Sard. To state it, a few definitions will be needed.

An *n-rectangle* $C \subset R^n$ is a nonempty set of the form $C = \{x : a_i \leqslant x_i \leqslant b_i, i = 1, \ldots, n\}$. The *measure* of C is $\prod_{i=1}^n (b_i - a_i)$. A subset $S \subset R^n$ has *measure zero* if for every $\epsilon > 0$ there exists a countable family of n-rectangles covering S, the sum of whose measures is less than ϵ.

Let $\phi : R^s \to R^t$ be a differentiable map. For any $x \in R^s$, $d\phi(x)$ will denote the Jacobian of ϕ at x. If range $(d\phi(x)) \neq R^t$, then x is a *critical point* for ϕ. If $y = \phi(x)$ for some critical point x, then y is a *critical value* for ϕ; otherwise, y is a *regular value* for ϕ.

SARD'S THEOREM. *Let $\phi : R^s \to R^t$ be differentiable of class C^r, where $r > \max(0, s - t)$. Then the set of critical values for ϕ is of measure zero in R^t.*

For a proof of Sard's Theorem, consult Abraham and Robbin [1].

PROOF OF THEOREM 1. For $x \in R^n$ and $\emptyset \neq S \subset I$, let $g^S(x)$ denote the $|S|$-dimensional vector with components $g_i(x)$ $(i \in S)$. By Sard's Theorem, the set of critical values for g^S is of measure zero in $R^{|S|}$. So, if $u \mapsto u^S$ denotes projection onto $R^{|S|}$,

$$N(S) = \{ u \in R^m : u^S \text{ is a critical value for } g^S \}$$

is of measure zero in R^m (a subset of $R^m = R^{m-|S|} \times R^{|S|}$ whose $R^{|S|}$-sections are all of measure zero is itself of measure zero; this follows from Fubini's Theorem). Then $N = \bigcup_{\emptyset \neq S \subset I} N(S)$ is also of measure zero.

Suppose $u \notin N$ and $x \in R^n$, and let $S = I(u, x)$. Since $u \notin N(S)$, $\{\nabla g_i(x) : i \in I(u, x)\}$ is linearly independent. Hence (CQ) is satisfied at x.

3. Second-order conditions. When the functions f and g_i are twice differentiable, a vector x is said to satisfy the *strong second-order conditions* for local optimality in $(Q(v, u))$ if x is feasible, condition (CQ) holds for x, and there exists $y \in R^m_+$ such that (KT) holds with

$$y_i > 0 \quad \text{for all } i \in I(u, x), \quad \text{and} \tag{3.1}$$

$$\begin{aligned} &\text{every nonzero } z \in R^n \text{ for which } z \cdot \nabla g_i(x) = 0 \\ &\text{for all } i \in I(u, x) \text{ satisfies } z \cdot H(x, y)z > 0, \end{aligned} \tag{3.2}$$

where $H(x, y)$ is the Hessian of the Lagrangian function in $(Q(v, u))$:

$$H(x, y) = \nabla^2 f(x) + \sum_I y_i \nabla^2 g_i(x).$$

These conditions are known to guarantee that x is an isolated locally optimal solution to $(Q(v, u))$. They also have other important consequences, for example with respect to the sensitivity of x to changes in value of the components of v and u; cf. Hestenes [2], McCormick [3]. (Somewhat weaker conditions would be enough to guarantee local optimality.)

THEOREM 2. *Suppose the functions $f, g_1, \ldots, g_m$ are of class C^2. Then except for (v, u) belonging to a set of measure zero in $R^n \times R^m$, $(Q(v, u))$ is such that for every locally optimal solution x and every $y \in R^m_+$ satisfying the first-order conditions (KT) with x, the strong second-order conditions actually hold.*

Before passing to the proof, we note that here, in contrast to theorem 1, the parameter vector v has an essential role. Generally speaking, there is no assurance that for a *given* v the conclusion will hold for almost every u, and indeed this can be false. This serves to emphasize how crucial it is to have the right concept of the "class" of problems relative to which a generic statement is to be made. Too large a class may render the result meaningless (the actual problems of interest may themselves form a negligible subclass, because of their special structure). But too small a class may render the result invalid. We refer again to Spingarn [4] for a more refined theory of the kinds of parameterizations that are workable.

COROLLARY. *Suppose f is of class C^2 and $g_1, \ldots, g_m$ are of class C^n. Then except for (v, u) belonging to a set of measure zero in $R^n \times R^m$, $(Q(v, u))$ is such that every locally optimal solution x satisfies the strong second-order conditions.*

The corollary is obtained by combining theorems 1 and 2. The exceptional u-set in theorem 1 is of measure zero for each v, so the exceptional (v, u)-set in $R^n \times R^m$ is also of measure zero.

PROOF OF THEOREM 2. For $t \in R$, let $\theta(t) = \max^2(0, t)$. Then θ is of class C^1 with $\theta'(t) = 2 \cdot \max(0, t)$. Consider the map $\Psi : R^n \times R^m \to R^n \times R^m$ defined for any $x' \in R^n$ and $\beta = (\beta_1, \ldots, \beta_m) \in R^m$ by

$$\Psi(x', \beta) = \begin{bmatrix} \nabla f(x') + \sum_I \theta(-\beta_i) \nabla g_i(x') \\ g_1(x') + \theta(\beta_1) \\ \vdots \\ g_m(x') + \theta(\beta_m) \end{bmatrix}$$

This map is of class C^1, so by Sard's Theorem its set of critical values is of measure zero in $R^n \times R^m$.

Let (v, u) be a regular value for Ψ and suppose x to be a local minimizer for $(Q(v, u))$. Let $y \in R^m_+$ be such that (KT) is satisfied. From the definition of Ψ, there is a unique $\beta \in R^m$ such that

$$\Psi(x, \beta) = \begin{pmatrix} v \\ u \end{pmatrix}, \tag{3.3}$$

$$\theta(-\beta_i) = y_i \quad \text{for all } i \in I, \text{ with } \beta_i \leqslant 0 \text{ iff } i \in I(u, x). \tag{3.4}$$

Let $G(x)$ denote the $n \times m$ matrix whose columns are the gradients $\nabla g_i(x)$, and let

$$D(\beta) = \begin{bmatrix} \theta'(\beta_1) & & \\ & \ddots & \\ & & \theta'(\beta_m) \end{bmatrix}.$$

The Jacobian of Ψ at (x, β) is

$$d\Psi(x, \beta) = \left(\begin{array}{c|c} H(x, y) & -G(x)D(-\beta) \\ \hline G(x)^T & D(\beta) \end{array} \right).$$

Since (v, u) is a regular value for Ψ, it follows from (3.3) that $d\Psi(x, \beta)$ is nonsingular. If for some $i \in I$ we had $\beta_i = 0$, then the $(n + i)$th column of $d\Psi(x, \beta)$ would vanish, in contradiction to nonsingularity. Hence by (3.4),

$$\beta_i < 0 \Leftrightarrow i \in I(u, x) \Leftrightarrow y_i > 0. \tag{3.5}$$

In particular, part (3.1) of the strong second-order conditions holds. Furthermore, since for $\beta_i < 0$ one has $\theta'(\beta_i) = 0$ and $\theta'(-\beta_i) > 0$, we may conclude from (3.5) and the nonsingularity of $d\Psi(x, \beta)$ that the gradients $\nabla g_i(x)$ for $i \in I(u, x)$ are linearly independent. Thus condition (CQ) holds.

It remains only to show that the Hessian condition (3.2) is satisfied. Since x is locally optimal and the constraint qualification (CQ) holds, we know from the theory of necessary conditions for optimality that at least $z \cdot H(x, y)z \geqslant 0$ is true in (3.2) (cf. Hestenes [2]). Therefore, if (3.2) were violated by some z, it would have to be with $z \cdot H(x, y)z = 0$, and z would thus be a *nonzero* optimal solution to the problem of minimizing $h(z') = \frac{1}{2} z' \cdot H(x, y)z'$ subject to $z' \cdot \nabla g_i(x) = 0$ for all $i \in I(u, x)$. Here $\nabla h(z) = H(x, y)z$ (because $H(x, y)$ is symmetric), so the theory of first-order optimality conditions implies the existence of scalars w_i for $i \in I(u, x)$ such that

$$H(x, y)z + \sum_{I(u, x)} w_i \nabla g_i(x) = 0. \tag{3.6}$$

Recalling that for $i \in I(u, x)$ we have $\theta'(\beta_i) = 0$ and $\theta'(-\beta_i) > 0$, while for $i \notin I(u, x)$ we have $\theta'(\beta_i) > 0$ and $\theta'(-\beta_i) = 0$, we define

$$s_i = \begin{cases} -w_i/\theta'(-\beta_i) & \text{if } i \in I(u, x), \\ -(\nabla g_i(x) \cdot z)/\theta'(\beta_i) & \text{if } i \notin I(u, x). \end{cases}$$

Then

$$d\Psi(x, \beta) \cdot \begin{pmatrix} z \\ s \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

as follows for the top half of the array from (3.6) and for the bottom half from the constraint $z \cdot \nabla g_i(x) = 0$ for rows $i \in I(u, x)$ and from the choice of s_i for rows

$i \notin I(u, x)$. Since $d\Psi(x, \beta)$ is nonsingular, we must have $z = 0$, $s = 0$. But z was supposed to be nonzero. This contradiction validates (3.2) and finishes the proof of theorem 2.

References

[1] Abraham, R. and Robbin, J. (1967). *Transversal Mappings and Flows*. W. H. Benjamin, Inc.
[2] Hestenes, M. R. (1975). *Optimization Theory: The Finite Dimensional Case*. John Wiley and Sons.
[3] McCormick, G. (1976). Optimality Criteria in Nonlinear Programming. *SIAM-AMS Proceedings*. **9** 27–38.
[4] Spingarn, J. E. (1977). Generic Conditions for Optimality in Nonlinear Programming Problems. Doctoral dissertation, University of Washington.

SPINGARN: DEPARTMENT OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GEORGIA 30332
ROCKAFELLAR: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WASHINGTON 98195