MATHEMATICS OF OPERATIONS RESEARCH
Vol. 20, No. 4, November 1995
Printed in U.S.A.

SECOND-ORDER SUFFICIENCY AND QUADRATIC GROWTH FOR NONISOLATED MINIMA

JOSEPH FRÉDÉRIC BONNANS AND ALEXANDER IOFFE

For standard nonlinear programming problems, the weak second-order sufficient condition is equivalent to the quadratic growth condition as far as the set of minima consists of isolated points and some qualification hypothesis holds. This kind of condition is instrumental in the study of numerical algorithms and sensitivity analysis. The aim of the paper is to study the relations between various types of sufficient conditions and quadratic growth in cases when the set of minima may have nonisolated points.

1. Introduction. The paper is devoted to the theory of nonlinear programming problems, i.e., finite-dimensional optimization with a finite number of constraints. The importance of second-order sufficient conditions is largely determined by their role in sensitivity analysis and numerical optimization. More attentive analysis of existing proofs (e.g., Bonnans (1992), Bonnans, Ioffe and Shapiro (1980, 1992), Gauvin and Janin (1988), Ioffe (1994)) shows, however, that, at least as far as sensitivity analysis is concerned, what is needed is not a second-order sufficient condition as such but rather an estimate of the kind (Shapiro (1988, 1992))

$$f(x) \geq c + \beta \operatorname{dist}^2(S, x), \tag{1.1}$$

in which f is the cost function, S is a set on which f has constant value c, and β a positive parameter. When S is a singleton, or more generally a finite set, the standard second-order sufficient condition (e.g., Ben-Tal and Zowe (1982), Ioffe (1979b), Levitin, Miljutin and Osmolowski (1974)) is equivalent to (1.1) provided the Mangasarian-Fromovitz constrained qualification is valid (e.g., Bonnans (1992)). But very little has been known so far about sufficient conditions and (1.1) like estimates in situations when the set of solutions has a more complicated structure than just a finite collection of isolated points.

This paper is an attempt to fill the gap. (The follow-up article (Bonnans and Ioffe (1995)) devoted to convex problems contains a complete characterization in that specific case.) We establish several sufficient conditions, based on second-order information, critical cones and proximal normals to the solution set at different levels of generality and simplicity of formulations which imply a general "quadratic growth condition" similar to (1.1). The formulation of the most general of them—we call it the "general sufficient condition" in the paper—seems to be fairly awkward at first glance. It requires information which is not "intrinsic" in the sense that it relies upon the existence of a certain "projection" map to the solution set with some special properties. (Although the proofs provide information on possible structure of the map, we cannot offer much practical advice for its construction.) What makes us

Received April 16, 1993; revised July 5, 1994
AMS 1991 subject classification. Primary: 90C30.
OR/MS Index 1978 subject classification. Primary: 659 Programming/Nonlinear/Theory/Optimality conditions.
Key words. Optimality conditions, Lagrangian function, composite functions, contingent cone, proximal normal, critical cone.

0364-765X/95/2004/0801/$01.25

introduce this condition as the basic sufficiency statement is that it is equivalent to the general growth condition under an additional "tangency" assumption which has a simple and natural formulation.

Considerations involving contingent cones are also instrumental in describing a (fairly general) structure of solution sets for which a sufficient condition very close to the standard second-order sufficient condition can be formulated. They also help to highlight the "bottle-neck" at which all the main difficulties caused by nonunicity of solutions are accumulated, namely the critical directions close to the contingent cone to the set of solutions. Much effort has been spent in the article to investigate the behaviour of the problem near such directions. Still some interesting questions remain unsolved.

A big portion of the paper is devoted to discussions on unconstrained optimization of a simple composite function (maximum of a finite collection of smooth functions) and only at the final section do we reformulate all the main results for constrained optimization problems using some simple reduction arguments. An advantage of such an approach (already tested for necessary conditions (Ioffe (1979b)) and sensitivity analysis (Ioffe (1994)) is that it allows us to get rid of feasibility problems in the course of main arguments.

2. Main results.

We begin by considering the function

$$f(x) := \max_{1 \le i \le m} f_i(x).$$

The functions f_i *are assumed to be twice continuously differentiable* from $\mathbb{R}^q$ into $\mathbb{R}$ throughout the paper. We use the following notation and terminology:

$$I(x) := \{i;\ 1 \le i \le m,\ f_i(x) = f(x)\},$$

the set of *active* indices,

$$\mathscr{L}(\lambda, x) := \sum_{i=1}^{m} \lambda_i f_i(x),$$

the *Lagrangian* of f,

$$\mathscr{S}^m := \left\{\lambda \in \mathbb{R}^m;\ \lambda \ge 0,\ \sum_{i=1}^{m} \lambda_i = 1\right\},$$

the standard simplex of $\mathbb{R}^m$,

$$\Omega(x) := \left\{\lambda \in \mathscr{S}^m;\ \lambda_i \ge 0,\ \lambda_i = 0 \text{ if } i \notin I(x);\ \sum_{i=1}^{m} \lambda_i \nabla f_i(x) = 0\right\}$$

(where as usual $\nabla f_i(x)$ is the gradient of f_i at x) the set of *Lagrange multipliers* for f at x and

$$\Omega_\delta(x) := \left\{\lambda \in \mathscr{S}^m,\ \lambda_i = 0 \text{ if } i \notin I(x);\ \left\|\sum_{i=1}^{m} \lambda_i \nabla f_i(x)\right\| \le \delta\right\},$$

the set of *Lagrange δ-multipliers*.

We call a point x *stationary* if $\Omega(x) \neq \emptyset$ and δ-stationary if $\Omega_\delta(x) \neq \emptyset$. We set further

$$C(x) := \{h: \nabla f_i(x)h \leq 0, \forall i \in I(x)\},$$

the cone of critical vectors of f at x.

In what follows *we fix a compact set S of stationary points* of f such that $f(x) \equiv \text{const} = c_0$ and S.

DEFINITION 1. A mapping π from a neighborhood U of S onto S will be called a *regular projection* onto S if there exists $\epsilon > 0$ such that

$$\epsilon\|x - \pi(x)\| \leq \operatorname{dist}(S, x), \qquad x \in U.$$

In particular, $\pi(x) = x$ for all $x \in S$.

Given a set $C \subset \mathbb{R}^q$ and $x \in C$, we denote by $T_C(x)$ the contingent cone to C at x:

$$T_C(x) := \limsup_{t \searrow 0} t^{-1}(C - x).$$

DEFINITION 2. Let C, D be sets and $x \in C \cap D$. We say that C and D are *nontangent* at x if

$$T_C(x) \cap T_D(x) = \{0\}.$$

DEFINITION 3. We say that a closed set $C \subset \mathbb{R}^k$ is *nice* if for every $x \in C$ there is a neighborhood U of x and a diffeomorphism F of U into $\mathbb{R}^k$ such that $C \cap U$ can be represented as a union of a finite number of (relatively closed) sets C_i which are nontangent to each other at x and such that the sets $F(C_i)$ are convex. We shall call the C_i *components* of C at x.

We say that f satisfies the *quadratic growth condition* (*QGC*) *on* S if:

(2.2) there exists $\beta > 0$ and a neighborhood U of S such that

$$f(x) \geq c_0 + \beta \operatorname{dist}^2(S, x), \qquad \forall x \in U.$$

We say that f satisfies the *general second-order sufficient condition* (*GSO*) on S if

for any $\delta > 0$ there exist a neighborhood U of S,

(2.3) a regular projection $\pi: U \to S$ and $\alpha > 0$ such that, for all $x \in U \setminus S$,

$$\max_{\lambda \in \Omega_\delta(\pi(x))} \left[\mathcal{L}_x(\lambda, \pi(x))h + \tfrac{1}{2}\mathcal{L}_{xx}(\lambda, \pi(x))(h, h)\right] \geq \alpha\|h\|^2,$$

where $h = x - \pi(x)$.

We also say that f satisfies the *tangency condition* (*TC*) on $D \subset \mathbb{R}^q$ if for any x in D,

for any $i \in I(x)$ either $i \in I(y)$ for all $y \in D$ sufficiently close to x,

or D and $\{y: f_i(y) = f_i(x) = c_0\}$ are nontangent at x.

If not specified, the set D is taken equal to S.

THEOREM 1. *The following implications hold*:

$$(GSO) \Rightarrow (QGC), \qquad (QGC) \,\&\, (TC) \Rightarrow (GSO).$$

THEOREM 2. *Suppose that*
(i) *S is a nice compact set of stationary points of f on which f is constant*,
(ii) *f satisfies (TC) on every component of S*,
(iii) *for any $x \in S$ and any $h \in C(x) \setminus T_S(x)$*,

$$\liminf_{u \xrightarrow{S} x} \max_{\lambda \in \Omega(u)} \mathscr{L}_{xx}(\lambda, u)(h, h) > 0. \tag{2.4}$$

Then (GSO) holds.

THEOREM 3. *If (QGC) holds, then*

$$\max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) \geq \beta \operatorname{dist}^2(T_S(x), h), \qquad \forall h \in C(x), \forall x \in S,$$

β being the constant defined in (2.1). *In particular*,

$$\max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) > 0, \qquad \forall x \in S, \forall h \in C(x) \setminus T_S(x). \tag{2.5}$$

2.1. Comments and corollaries.

2.4.1. Strictly speaking, (GSO) is not a second-order condition. It holds, for instance, for piecewise linear functions (or, equivalently, for linear program) in which case we actually have a stronger "linear growth condition" (Ioffe (1994)). A "pure" second-order condition that we can distill from Theorem 1 is the following.

COROLLARY 1. *Suppose that the following property holds*:
(GSO_1) *there exist $\alpha, \beta > 0$, a neighborhood U of S and a regular projection π: $U \to S$ such that for $h := x - \pi(x)$, we have*

$$\tfrac{1}{2} \max_{\lambda \in \Omega(\pi(x))} \mathscr{L}_{xx}(\lambda, \pi(x))(h, h) \geq \alpha \|h\|^2$$

whenever $x \in U$ satisfies $f(x) \leq c_0 + \beta \operatorname{dist}^2(S, x)$.
Then (QGC) holds.

PROOF. We may assume $\beta < \alpha$. We observe that the proof of Theorem 1 actually shows that, reducing if necessary the neighbourhood U, the implication (2.3) $\Rightarrow$ (2.2) always holds for any given x. Therefore if

$$f(x) \leq c_0 + \beta \operatorname{dist}^2(S, x)$$

(otherwise (2.2) is trivial), then, as every point of S is stationary and $\Omega(y) \subset \Omega_\delta(y)$,

$$\max_{\lambda \in \Omega_\delta(x)} \left\{ \mathscr{L}_x(\lambda, \pi(x))h + \tfrac{1}{2}\mathscr{L}_{xx}(\lambda, \pi(x))(h, h) \right\}$$

$$\geq \tfrac{1}{2} \max_{\lambda \in \Omega(\pi(x))} \mathscr{L}_{xx}(\lambda, \pi(x))(h, h) \geq \alpha \|h\|^2,$$

which is (2.3). □

2.4.2. The main advantage of Theorems 2 and 3 over Theorem 1 is that they are intrinsic, i.e., are stated in terms of the original data only, while Theorem 1 requires a foreign object such as a "regular projection." Further intrinsic sufficient criteria, easier to verify than that of Theorem 2, can be found in §4.

Here we only observe that the standard second-order sufficient condition is an easy corollary of Theorem 2, for the conditions (i) and (ii) of Theorem 2 are automatically satisfied if S is a finite set, as then $T_S(x) = \{0\}$ for any $x \in S$. On the other hand, if S is finite and (QGC) is satisfied, then any $x \in S$ is a local minimum of the function $f(x+h) - \alpha\|h\|^2$ for some $\alpha > 0$: applying the second-order necessary condition, we arrive at the following local characterization of quadratic growth for isolated minima.

COROLLARY 2. *Let S be a finite set of stationary points of f. Then f satisfies (QGC) on S if and only if*

$$\max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) > 0, \qquad \forall x \in S, \forall h \in C(x),\ h \neq 0.$$

2.4.3. The proof of Theorem 2 in the next section actually shows that the conclusion of the theorem remains valid if we replace (iii) by the refined condition below:

(iii′) condition (2.5) holds and there exists $\epsilon > 0$ such that (2.4) is valid for all $h \in C(x) \cap (T_S^\epsilon(x) \setminus T_S(x))$, where $T_S^\epsilon(x)$ is the set of approximate contingent directions to S at x:

$$T_S^\epsilon(x) = \{h \in \mathbb{R}^n : \operatorname{dist}(T_S(x), h) \le \epsilon\|h\|\}.$$

We observe further that (2.5) is actually necessary for (QGC) to hold as follows from Theorem 3. It is therefore natural to ask whether it is possible to get rid of (iii) or (iii′) altogether and to replace it by (2.5) in Theorem 2. The following example shows that (iii) cannot be a necessary condition for (QGC) even in its modified (iii′) form.

Let $X = \mathbb{R}^2$, $x = (x_1, x_2)$, and (see the figure):

$$f(x) := \max\{-x_1x_2 + x_2^2,\ x_1x_2 - 2x_2^2 - x_1,\ 2x_2 - x_1,\ x_1 - 1\};$$

$$S := \{x = (x_1, x_2) : 0 \le x_1 \le 1,\ x_2 = 0 \text{ or } x_2 = x_1/2\}.$$

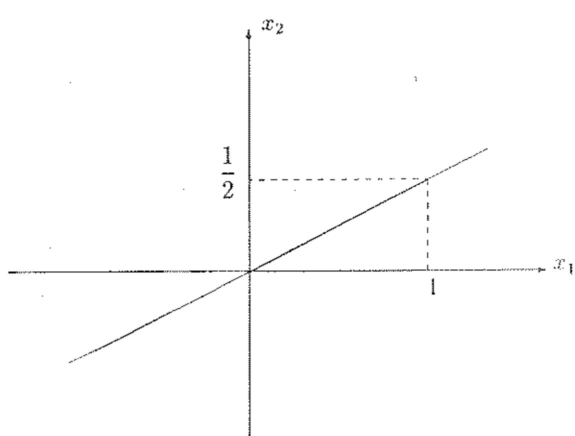

FIGURE 1.

It can be easily verified that f satisfies (QGC) on S. Indeed:
if $x_2 \le 0$, $0 \le x_1 \le 1$, then $f(x) \ge x_2^2 = \operatorname{dist}^2(S, x)$;
if $x_1 \le 0$ then

$$f(x) \ge \max\{-x_1, x_2^2 - x_2x_1\} \ge \max\{x_1^2, |x_2|(|x_2| - |x_1|)\}$$
$$\ge \tfrac{1}{2}\left[x_1^2 + \left(x_2^2 - |x_1x_2|\right)\right] \ge \tfrac{1}{4}\left(x_1^2 + x_2^2\right) \ge \tfrac{1}{4}\|x\|^2;$$

If $0 \le x_2 \le x_1/2$, $0 \le x_1 \le 1$, then

$$f(x) \ge 2x_2\left(\frac{x_1}{2} - x_2\right) \ge 2\min\left\{x_2^2, \left(\frac{x_1}{2} - x_2\right)^2\right\} \ge 2\operatorname{dist}^2(S, x),$$

etc.

We notice furthermore that $T_S(x) = C(x)$ at any $x \in S$, $x \neq 0$ whereas

$$T_S(0) = \{h = (h_1, h_2) : h_1 \ge 0,\ h_2 = 0 \text{ or } h_2 = h_1/2\},$$

$$C(0) = \{h = (h_1, h_2) : h_1 \ge 0,\ h_2 \le h_1/2\}.$$

If $x = (x_1, 0) \in S$, $x_1 > 0$, then $I(x) = \{1, 2\}$ and $\mathscr{L}_x(x) = -\lambda_1 x_1 + \lambda_2 x_1 = 0$ which implies $\lambda_1 = \lambda_2 = 1/2$. Therefore for any $h = (h_1, h_2)$,

$$\max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) = -h_2^2/2.$$

Now taking $h \in C(0) \setminus T_S(0)$ which means that $h_2 \neq 0$ (and $h_2 < h_1/2$) we see that

$$\liminf_{\substack{x \to 0 \\ S}} \max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) \le -\tfrac{1}{2}h_2^2 < 0.$$

Hence (iii) or (iii′) are not satisfied, as was to be proved.

On the other hand condition (2.5) alone is not sufficient for (QGC), even if S is smooth. Indeed, consider the function

$$f(x) = \max\left(x_1 x_x^2, -x_1, -x_2, x_2 - x_1^2, 1 - x_1\right).$$

Then the minimum value 0 is attained on $S = [0, 1] \times \{0\}$. It happens that the set of critical directions is equal to the contingent set of S at all x in S, so that (2.5) is trivially satisfied. However $x(t) := (t, t)$ with $t > 0, t \to 0$ satisfies $f(x(t)) = t^3$ and $\operatorname{dist}(x(t), S) = t$, hence (QGC) does not hold. We note that Theorem 2 excludes this case as (TC) is not satisfied over S.

2.4.4. It is useful to have simple verifiable criteria for conditions (iii) or (iii′). Thanks to Theorem 3, (iii) is satisfied if (2.5) holds and $\lim \Omega(u)$ as $u \to x$ in S' exists and is equal to $\Omega(x)$ (we take the limit, lim sup, lim inf of sets in the sense of Painlevé-Kuratowski). However, in general, $\Omega(\cdot)$ is no more than upper semicontinuous and the continuity requirement is rather strong.

Two obvious (though important) cases when it is satisfied could be mentioned: when gradients of active f_i are linearly independent at x, and when the functions are convex.

In connection with this question we would like to draw attention to the following elementary fact.

PROPOSITION 1. *We assume that (TC) holds on S and let*

$$I_0(x) = \liminf_{u \xrightarrow{S} x} I(x).$$

Then for any $\lambda \in \Omega(x)$ and $i \in I(x) \setminus I_0(x)$ we have $\lambda_i \nabla f_i(x) = 0$.

PROOF. Pick $i \in I(x) \setminus I_0(x)$. There exists a sequence of $x_n \in S$, $x_n \to x$ such that $i \notin I(x^n)$. Extracting if necessary a subsequence, we may assume that $h = \lim \|x^n - x\|^{-1}(x^n - x)$. Then

$$\nabla f_i(x)h = \lim_{n\to\infty} \|x^k - x\|^{-1}(f_i(x^n) - f_i(x))$$
$$\leq \lim_{n\to\infty} \|x^k - x\|^{-1}(f(x^n) - f(x)) = 0.$$

On the other hand, if $\nabla f_i(x) \neq 0$, then the equality $\nabla f_i(x)h = 0$ (meaning that h is in the tangent cone to the level set of f) is by (TC) impossible as $h \in T_S(x)$ by definition. Hence $\nabla f_i(x)h < 0$. If $\lambda \in \Omega(x)$, then $0 = \sum_{i=1}^m \lambda_i \nabla f_i(x) \cdot d$ and each term of the sum is not positive, hence null. This implies $\lambda_i = 0$. □

We note that if $f_i(x)$, $i = 1, \ldots, m$ are convex, then the set of stationary points of f is actually convex and equal to the set of minima of f so that hypothesis (i) of Theorem 2 is satisfied.

3. Proofs of the theorems. We first prove Theorem 1. We need a preliminary lemma.

LEMMA 1. *Assume that (TC) holds on a set $D \subset S$. Given two sequences, $\{u^n\} \subset D$ and $\{x^n\} \subset \mathbb{R}^q$, both converging to $x \in D$, and such that $\|u^n - x^n\| = o(\|u^n - x\|)$ and $f(x^n) \geq f(x)$. Then $I(x^n) \subset I(u^n)$ for all sufficiently large n.*

PROOF. Assume the contrary. Taking if necessary a subsequence, we reduce to the situation when $I(u^n) = J$ and there exists an $i \in I(x^n) \setminus J$. Therefore, $f_i(x^n) = f(x^n) \geq f(x) = f(u^n) > f_i(u^n)$. It follows that for some $y^n \in [x^n, u^n]$ we have $f_i(y^n) = f(u^n) = f(x)$.

We may assume the existence of $h := \lim(u^n - x)/\|u^n - x\|$. Obviously, $h \in T_D(x)$. On the other hand, $h = \lim(y^n - x)/\|y^n - x\|$, so that h belongs to the contingent cone to $\{y\colon f_i(y) = f_i(x)\}$, in contradiction with (TC). □

PROOF OF THEOREM 1. (GSO) ⇒ (QGC). Fix $\delta > 0$, a neighbourhood U of S and a regular projection $\pi\colon U \to S$ such that (2.3) holds. Choose $0 < \beta < \alpha$ and a $\sigma > 0$ small enough to make sure that

$$\left|\mathscr{L}(\lambda, x+h) - \mathscr{L}(\lambda, x) - \mathscr{L}_x(\lambda, x)h - \tfrac{1}{2}\mathscr{L}_{xx}(\lambda, x)(h, h)\right| < (\alpha - \beta)\|h\|^2,$$

provided $x \in S$, $\lambda \in S^m$ and $\|h\| < \sigma$. Pick $x \in U$. With no loss of generality we may assume that $\|x - \pi(x)\| < \sigma$. Setting $h = x - \pi(x)$, we obtain

$$\begin{aligned} f(x) - c_0 &= f(\pi(x) + h) - f(\pi(x)), \\ &\geq \max_{\lambda \in \Omega_\delta(\pi(x))} \{\mathscr{L}(\lambda, \pi(x) + h) - \mathscr{L}(\lambda, \pi(x))\}, \\ &\geq \max_{\lambda \in \Omega_\delta(\pi(x))} \{\mathscr{L}_x(\lambda, \pi(x))h + \tfrac{1}{2}\mathscr{L}_{xx}(\lambda, \pi(x))(h, h) - (\alpha - \beta)\|h\|^2\}, \\ &\geq \beta\|h\|^2 \geq \beta \operatorname{dist}(S, x)^2, \end{aligned}$$

i.e., (QGC) holds.

(QGC) & (TC) ⇒ (GSO). Assume the contrary. As S is compact there exist $\delta > 0$ and $x^n \xrightarrow{S} x$ such that if a sequence $u^n \in S$ satisfies $\|u^n - x^n\| \le n \operatorname{dist}(S, x^n)$ we have

$$\max_{\lambda \in \Omega_\delta(u^n)} \left(\mathscr{L}_x(\lambda, u^n) h^n + \tfrac{1}{2} \mathscr{L}_{xx}(\lambda, u^n)(h^n, h^n) \right) \le \tfrac{1}{n} \|h^n\|^2, \tag{3.6}$$

where $h^n = x^n - u^n$. We next choose u^n by considering the following two cases:

(A) $\operatorname{dist}(S, x^n) = o(\|x - x^n\|)$ for a subsequence that we may assume to be equal to the sequence itself. Choose $u^n \in S$ such that $\|x^n - u^n\| = \operatorname{dist}(S, x^n)$, and set $h^n := x^n - u^n$. From (QGC), it follows that $f(x^n) \ge f(x)$. Applying Lemma 1, we deduce that $I(x^n) \subset I(u^n)$ whenever n is large enough.

(B) If (A) is not satisfied, then there exists $\theta > 0$ such that

$$\operatorname{dist}(S, x^n) \ge \theta \|x^n - x\|,$$

and then we choose $u^n = x$. As $I(u)$ is an upper semicontinuous map, we have $I(x^n) \subset I(x)$ for large n.

In both cases, by (QGC) we have

$$\begin{aligned} \beta \|h^n\|^2 &\le f(u^n + h^n) - f(u^n) \\ &= \max_{i \in I(u^n)} \{ f_i(u^n + h^n) - f_i(u^n) \} \\ &= \max_{\lambda \in \Omega_x(u^n)} \left\{ \sum_{i=1}^m \lambda_i (f_i(u^n + h^n) - f_i(u^n)) \right\} \\ &= \max_{\lambda \in \Omega_x(u^n)} \{ \mathscr{L}_x(\lambda, u^n) h^n + \tfrac{1}{2} \mathscr{L}_{xx}(\lambda, u^n)(h^n, h^n) \} + o(\|h^n\|^2), \end{aligned} \tag{3.7}$$

with $o(\|h^n\|)/\|h^n\| \to 0$ uniformly over $\lambda \in \Omega_x(u^n)$. Set

$$\xi := \max\{ \|\mathscr{L}_x(\lambda, x)\|;\ x \in S,\ \lambda \in \Omega_x(x) \}.$$

Note that $\Omega_x(x) = \{\lambda \in \mathscr{S}^m;\ \lambda_i = 0 \text{ if } i \notin I(x)\}$. We have

$$\delta \xi^{-1} \Omega_x(x) \subset \Omega_\delta(x),$$

so with (3.6) and (3.7),

$$\begin{aligned} \beta \|h^n\|^2 &\le \frac{\xi}{\delta} \max_{\lambda \in \Omega_\delta(u^n)} \left(\mathscr{L}_x(\lambda, u^n) h^n + \frac{1}{2} \mathscr{L}_{xx}(\lambda, u^n)(h^n, h^n) \right) + o(\|h^n\|^2) \\ &\le \frac{\xi}{n\delta} \|h^n\|^2 + o(\|h^n\|^2) = o(\|h^n\|^2), \end{aligned} \tag{3.8}$$

a contradiction to (QGC).

Proof of Theorem 2. The proof is based on the following lemma, to be proved a bit later.

Lemma 2. *Under the hypotheses of Theorem 2, if $x^n \to x \in S$, $I(x^n) = J$ and $\operatorname{dist}(x^n, S) = o(\|x - x^n\|)$, then there exists $\{w^n\} \subset S$ such that*

$$\operatorname{dist}(S, x^n) = O(\|x^n - w^n\|),$$

and $e^n := \|x^n - w^n\|^{-1}(x^n - w^n)$ *have, among their limit points as* $n \to \infty$, *a vector* $e \notin T_S(x)$ *such that* $\nabla f_i(x)e \leq 0$ *for all* $i \in I(x) \setminus J$. *Moreover, given* $\epsilon > 0$, *the sequence of* w^n *can be chosen in such a way that* $\|h - e\| < \epsilon$, *where* $h \in T_S(x)$ *is a limit-point of* $\|x^n - x\|^{-1}(x^n - x)$.

Suppose the theorem is wrong and (GSO) is not valid. Then, as in the proof of Theorem 1, we find a $\delta > 0$ and a sequence of x^n converging to an $x \in S$ such that for any $u \in S$ with $\|u - x^n\| \leq n\,\mathrm{dist}(S, x^n)$, (3.6) holds.

Let $u^n \in S$ be a nearest to x^n, $t^n := \|x^n - x\|$, $h^n := t_n^{-1}(x^n - x)$, and let h^n converge to h, $\|h\| = 1$. We consider the same two possibilities as in 3.1.3 (but in the opposite order).

(A) $\|x - x^n\| \geq O(\mathrm{dist}(S, x^n))$. Then $h \notin T_S(x)$ and as $\|x - x^n\| \leq n\,\mathrm{dist}(S, x^n)$ for large n, so (3.6) must hold with h replaced by $x^n - x$ and u replaced by x. Therefore

$$\max_{\lambda \in \Omega_\delta(x)} \left[\mathscr{L}_x(\lambda, x)h^n + \frac{t_n}{2}\mathscr{L}_{xx}(\lambda, x)(h^n, h^n) \right] \leq \frac{t_n}{n}, \tag{3.9}$$

and, consequently, for any $\delta > 0$:

$$\max_{\lambda \in \Omega_0(x)} \mathscr{L}_x(\lambda, x)h \leq 0.$$

This may happen only if $h \in C(x)$. Thus $h \in C(x) \setminus T_S(x)$ and inequality (2.4) is valid for h and x, in particular

$$\max_{\lambda \in \Omega(x)} \mathscr{L}(\lambda, x)(h, h) > 0.$$

On the other hand it follows from (3.9) that

$$\max_{\lambda \in \Omega(x)} \frac{t_n}{2}\mathscr{L}_{xx}(\lambda, x)(h^n, h^n) \leq \frac{t_n}{n},$$

and therefore

$$\max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) \leq 0.$$

Hence we arrived at a contradiction.

(B) $\mathrm{dist}(S, x^n) = \|u^n - x^n\| = o(\|x - x^n\|)$. Then $h \in T_S(x)$.

Assume for the moment that the lemma has already been proved. Find a sequence of w^n as in Lemma 2 and let $e \notin T_S(x)$ be a corresponding limit point.

As $\|x^n - w^n\|$ is of the same order as $\mathrm{dist}(S, x^n)$, we have $\|x^n - w^n\| \leq n\,\mathrm{dist}(S, x^n)$ so that (3.6) holds with $x = w^n$ and $h = x^n - w^n$, that is to say

$$\max_{\lambda \in \Omega_\delta(w^n)} \left[\mathscr{L}_x(\lambda, w^n)e^n + \frac{\tau_n}{2}\mathscr{L}_{xx}(\lambda, w^n)(e^n, e^n) \right] \leq \frac{\tau_n}{n}, \tag{3.10}$$

where $\tau_n = \|x^n - w^n\|$.

We observe further that $\|w^n - u^n\|$ is both $o(\|w^n - x\|)$ and $o(\|u^n - x\|)$, so by Lemma 1, $I(w^n) = I(u^n) = J$ for large n.

It follows from (3.10) that

$$\nabla f_i(x)e = \lim \nabla f_i(x^n)e^n = 0 \qquad \forall i \in J,$$

and, by Lemma 2, $\nabla f_i(x)e \le 0$ if $i \in I(x) \setminus J$. Consequently, $e \in C(x) \setminus T_S(x)$, and (iii) implies that

$$0 < \liminf_{n\to\infty} \max_{\lambda \in \Omega(w^n)} \mathscr{L}_{xx}(\lambda, w^n)(e^n, e^n),$$

in contradiction with (3.10). □

PROOF OF LEMMA 2. By (i) there is a finite collection of closed convex sets $C_1, \ldots, C_k$ (say, containing zero) and a diffeomorphism Q of a neighbourhood V of zero onto a neighbourhood U of x such that

$$S \cap U = Q(C \cap V); \quad \text{where } C = \bigcup C_j.$$

Let y^n and v^n be defined by

$$Q(y^n) = x^n, \qquad Q(v^n) = u^n.$$

Then $v^n \to 0$, $y^n \to 0$, $v^n \in C$,

$$\|x^n - u^n\| = O(\|y^n - v^n\|), \qquad \|x^n - x\| = O(\|v^n\|) \tag{3.11}$$

and the sets $S_j = Q(C_j)$ are nontangent at x.

We may assume that all v^n belong to the same C_j say to C_1. We deduce that $u^n \in S_1$, $h \in T_{S_1}(x)$, $h \notin T_{S_j}(x)$, $j = 2, \ldots, k$. By (ii), $\nabla f_i(x)h < 0$ if $i \in I(x) \setminus J$ and $\nabla f_i(x) \neq 0$. Therefore we can find $\gamma > 0$ such that $\|e - h\| < \gamma$ implies that $e \notin T_{S_j}(x)$, $j = 2, \ldots, k$ and $\nabla f_i(x)e \le 0$ if $i \in I(x) \setminus J$, $\nabla f_i(x) \neq 0$. We can assume that $\gamma < \epsilon$. Fix $M > 1 + 2\gamma^{-1}$ and let

$$\alpha^n := M\|x^n - u^n\|/\|x^n - x\|,$$

$$z^n := (1 - \alpha^n)v^n, \tag{3.12}$$

$$w^n := Q(z^n).$$

Then $\alpha^n \to 0$, $z^n \in C_1$ and $w^n \in S_1$. We further define e^n as in the statement by means of the w^n. As always, we assume that $e^n \to e$. We have to show that $e \notin T_S(x)$ and that $\|h - e\| \le \epsilon$. We have

$$w^n = Q((1 - \alpha^n)v^n) = Q(0) + Q'(0)(1 - \alpha^n)v^n + o(\|v^n\|),$$

and, on the other hand,

$$w^n = Q(v^n - \alpha^n v^n) = Q(v^n) + Q'(v^n)(-\alpha^n v^n) + o(\alpha^n\|v^n\|).$$

Multiplying the first equality by α^n, the second by $(1 - \alpha^n)$ and adding, we have

$$w^n = \alpha^n x + (1 - \alpha^n)u^n + [Q'(0) - Q'(v^n)]\alpha^n(1 - \alpha^n)v^n + o(\alpha^n\|v^n\|),$$

$$= \alpha^n x(1 - \alpha^n)u^n + o(\alpha^n\|v^n\|),$$

$$= \alpha^n x + (1 - \alpha^n)u^n + o(\|u^n - x^n\|),$$

or

$$w^n - x^n = \alpha^n(x - x^n) + (1 - \alpha^n)(u^n - x^n) + o(\|u^n - x^n\|). \tag{3.13}$$

It follows from (3.12) and (3.13) that

$$\left| \frac{\|w^n - x^n\|}{\|u^n - x^n\|} - M \right| \le 1 + r^n, \tag{3.14}$$

where $r^n \to 0$. In particular, $\|w^n - x^n\| = O(\mathrm{dist}(S_1, x^n))$ from which, using the fact that S_1 is diffeomorphic to a convex set, we conclude that $e \notin T_{S_1}(x)$.

Thanks to the choice of γ, all we have to show is that $\|h - e\| < \gamma$. We have from (3.13) setting $g^n = (x^n - u^n)/\|x^n - u^n\|$:

$$e^n = \alpha^n \frac{\|x^n - x\|}{\|x^n - w^n\|} h^n + \frac{\|x^n - u^n\|}{\|x^n - w^n\|} g^n + r^n,$$

where $\|r^n\| \to 0$ or (by (3.12)),

$$e^n = \frac{\|x^n - u^n\|}{\|x^n - w^n\|}(Mh^n + g^n) + r^n,$$

which together with (3.14) gives

$$\|e^n - h^n\| \le \frac{2}{M-1} + r^n,$$

that is (see the choice of M):

$$\|e - h\| \le \frac{2}{M-1} < \gamma. \quad \text{Q.E.D.}$$

We end this section by proving Theorem 3.

PROOF OF THEOREM 3. We have (see, e.g., Ioffe (1991), Corollary 5),

$$\liminf_{\substack{\sigma \to 0 \\ h' \to h}} \frac{f(x + \sigma h') - f(x)}{\sigma^2} = \max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h), \tag{3.15}$$

for any $x \in S$ and any $h \in C(x)$. Assume now that the (QGC) holds. According to the definition of $T_S(x)$,

$$\mathrm{dist}(S, x + \sigma h) \ge \sigma\, \mathrm{dist}(T_S(x), h) + o(\sigma);$$

hence

$$f(x + \sigma h) \ge f(x) + \beta\sigma^2\, \mathrm{dist}(T_S(x), h)^2 + o(\sigma^2),$$

which, together with (3.15), immediately implies the theorem. □

4. Further intrinsic sufficient conditions. The proof of Theorem 1 suggests that the orthogonal projection onto S has a special importance for (GSO). We shall obtain some simple intrinsic sufficient conditions using this idea. Recall that a vector h is called a proximal normal to S at $x \in S$ if

$$t\|h\| = \text{dist}(S, x + th)$$

for sufficiently small $t > 0$ (enough to require that there is at least one $t > 0$ with such property). We shall denote by $\text{PN}(S, x)$ the collection of proximal normals to S at x. It is always a convex cone.

We also denote by $C_\epsilon(x)$ the ϵ-critical cone for f at x:

$$C_\epsilon(x) = \{h : \nabla f_i(x)h \le \epsilon\|h\|,\ i \in I(x)\}.$$

LEMMA 3. *Let $x^n \in S$ and $h^n \to 0$ be such that for a certain $\delta > 0$,*

$$\max_{\lambda \in \Omega_\delta(x^n)} \left[\mathscr{L}_x(\lambda, x^n)h^n + \tfrac{1}{2}\mathscr{L}_{xx}(\lambda, x^n)(h^n, h^n)\right] \le O(\|h^n\|^2).$$

Then given $\epsilon > 0$, there exists n_0 such that $h^n \in C_\epsilon(x^n)$ whenever $n > n_0$.

PROOF. We already observed in §3 that $\Omega_\infty(x) \subset \delta\xi^{-1}\Omega_\delta(x)$ for some $\xi > 0$. It follows from the assumption that

$$\max_{\lambda \in \Omega_\infty(x^n)} \left[\mathscr{L}_x(\lambda, x^n)h^n + \tfrac{1}{2}\mathscr{L}_{xx}(\lambda, x^n)(h^n, h^n)\right] \le O(\|h^n\|^2);$$

hence

$$\max_{\lambda \in \Omega_\infty(x)} \mathscr{L}_x(\lambda, x^n)h^n \le O(\|h^n\|^2).$$

On the other hand, for any $u \in S$ any h and any $i \in I(x)$, $\nabla f_i(x)h \le \max_{\lambda \in \Omega_\infty(x)} \mathscr{L}_x(\lambda, x)h$. The conclusion follows. □

PROPOSITION 2. *Suppose that there exist $\epsilon > 0$ and $\alpha > 0$ such that*

$$\max_{\lambda \in \Omega(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) \ge \alpha\|h\|^2$$

for any $x \in S$ and any $h \in C_\epsilon(x) \cap PN(S, x)$. Then (GSO) holds.

PROOF. Let π be an orthogonal projection onto S, i.e., $\|x - \pi(x)\| = \text{dist}(S, x)$. We shall show that (GSO) holds with such a π. Assuming the contrary we shall conclude that for any $\delta > 0$ there exists $\{x^n\} \subset S$ and $h^n \to 0$ such that $\|h^n\| = \text{dist}(S, x^n + h^n)$ and (3.6) holds with $u = x^n + h^n$, $h = h^n$. By Lemma 3, $h^n \in C_\epsilon(x^n)$ if n is large enough and, by definition $h^n \in \text{PN}(S, x^n)$. So we get a contradiction as soon as $\alpha > n^{-1}$. □

Calculation of ϵ-critical vectors may present certain difficulties compared with calculation of "regular" critical vectors. The next proposition gives a sufficient criterium in terms of the latter. For any $x \in S$ and h we set

$$f'(x; h) = \max_{i \in I(x)} \nabla f_i(x)h,$$

which is the directional derivative of f at x. Then $h \notin C(x)$ if and only if $f'(x; h) > 0$. For such h we set

$$I_1(x; h) = \{i \in I(x),\ \nabla f_i(x)h = f'(x; h)\};$$

$$M(x; h) = \{\lambda : \in \mathscr{S}^m;\ \lambda_i = 0 \text{ if } i \notin I_1(x; h);\ \mathscr{L}_x(\lambda, x)h \ge 0\};$$

$$\mu(x; h) = \min\{\|\mathscr{L}_x(\lambda, x)\| : \lambda \in M(x; h)\}.$$

We also set

$$PN_\delta(S,x) = \{h\colon \mathrm{dist}(\mathrm{PN}(S,x),h) \le \delta\|h\|\}.$$

PROPOSITION 3. *Suppose that there exist* $\bar{\mu} > 0$ *such that*

$$\mu(x;h) \ge \bar{\mu}, \qquad \forall x \in S, \forall h \notin C(x),$$

and that there exist $\alpha > 0,\ \delta > 0$ *such that*

$$\max_{\lambda\in\Omega(x)} \mathscr{L}_{xx}(\lambda,x)(h,h) \ge \alpha\|h\|^2$$

for all $x \in S$, $h \in C(x) \cap PN_\delta(S,x)$. *Then (GSO) holds.*

PROOF. We will apply Proposition 2 in order to get the result. So, let h be in $C_\epsilon(x) \cap \mathrm{PN}(S,x)$. It follows from Ioffe (1979a) that

$$\mathrm{dist}(C(x),h) \le \bar{\mu}^{-1} f'(x;h)$$

(due to homogeneity of $f'(x;\cdot)$). Therefore

$$h \in C_\epsilon(x) \Rightarrow \mathrm{dist}(C(x);h) \le \frac{\epsilon}{\bar{\mu}}\|h\|. \tag{4.16}$$

Choose $\delta_1 \in (0,1/2)$ such that

$$|\mathscr{L}_{xx}(\lambda,x)(h,h) - \mathscr{L}_{xx}(\lambda,x)(h',h')| \le \alpha/2 \tag{4.17}$$

if $x \in S$, $\lambda_i \ge 0$, $\Sigma\lambda_i = 1$, $\|h\| = 1$, $\|h-h'\| \le \delta_1$. Let $\epsilon > 0$ be so small that

$$\frac{\epsilon}{\bar{\mu}} < \min\left\{\frac{\delta_1}{1+\delta_1}, \frac{\delta}{1+\delta}\right\}. \tag{4.18}$$

By (4.16) and (4.18) for any $h \in C_\epsilon(x) \cap \mathrm{PN}(S,x)$ there is an $e \in C(x)$ such that

$$\|h-e\| \le \frac{\epsilon}{\bar{\mu}}\|h\| \le \delta\|e\|. \tag{4.19}$$

This means that $e \in C(x) \cap \mathrm{PN}_\delta(S,x)$. Then by hypothesis

$$\alpha\|e\|^2 \le \max_{\lambda\in\Omega(x)} \mathscr{L}_{xx}(\lambda,x)(e,e),$$

and, as $\|h-e\| \le \delta_1\|e\|$ by (4.18), (4.17) implies that

$$\max_{\lambda\in\Omega(x)} \mathscr{L}_{xx}(\lambda,x)(h,h) \ge \max_{\lambda\in\Omega(x)} \mathscr{L}_{xx}(\lambda,x)(e,e) - \frac{\alpha}{2}\|h\|^2$$

$$\ge \alpha\big(|e|^2 - 1/2)\|h\|^2\big).$$

Taking ϵ small enough and using (4.18) we can minorize the right-hand side by, say, $(\alpha/4)\|h\|^2$. We now just have to apply Proposition 2. □

It can be observed that Propositions 1 and 2, though much simpler to formulate, are weaker results than Theorems 1 and 2. To see this, we can consider the function

$$f(x) = \max\{\xi\eta, -\xi, -\eta, \xi + \eta - 1\}$$

(where $x = (\xi, \eta) \in R^2$), and

$$S = \{x : \xi\eta = 0; 0 \leq \xi, \eta, \xi + \eta \leq 1\}.$$

It can be easily verified that the conditions of Theorems 1, 2 and even Corollary 1 are satisfied on this case but not the conditions of Proposition 1 and 2.

5. Problems with constraints.

5.1. General case. This section is essentially devoted to the reformulation of the main results for constrained nonlinear programs:

$$\text{(P)} \qquad \begin{array}{l} \text{minimize } f_0(x) \\ \text{subject to } f_i(x) \leq 0,\ i = 1, \ldots, k;\ f_i(x) = 0,\ i = k+1, \ldots, m. \end{array}$$

The very fact that theorems on a maximum function as considered above can be applied to (P) follows from the simple observation (cf. Ioffe (1994)):

PROPOSITION 4. *Let S be a closed set of feasible elements of* (P) *such that $f_0(x) \equiv$ cont $= c$ on S. Set*

$$f(x) = \max\{f_0(x) - c, f_1(x), \ldots, f_k(x), |f_{k+1}(x)|, \ldots, |f_m(x)|\}. \tag{5.20}$$

Then the following two properties are equivalent:

(a) *there is a neighbourhood U of S such that $f_0(x) > c$ for any $x \in U \setminus S$ which is feasible for* (P);

(b) *$f(x) > 0$ for any $x \in U \setminus S$.*

PROOF. The implication (b) $\Rightarrow$ (a) is obvious. Conversely, if (a) holds, then $f(x) \geq f_0(x) > 0$ for any feasible $x \in U \setminus S$. On the other hand if x is not feasible, then either $f(x) > 0$ for some $i = 1, \ldots, k$ or $|f_i(x)| > 0$ for some $i = k+1, \ldots, m$; in either case $f(x) > 0$. □

Thanks to this proposition we can easily reformulate the basic properties, i.e., the quadratic growth condition and the second-order sufficient condition, as well as all the theorems for (P), using the specific form of the function f given by (5.20).

The reformulation procedure actually consists of: (a) replacing $|f_i(x)|$ by $\max\{f_i(x), -f_i(x)\}$ in (5.20) followed by application of all the formulae to the so-obtained function and the subsequent return to the original notation, and, (b) the observation that $f(x)$ and $f_i(x)$ for $i = k+1, \ldots, m$ are constant on S.

The results of the reformulation can be summarized as follows. Consider the set $\Lambda(x)$ of Lagrange multipliers of (P) at x:

$$\Lambda(x) = \Big\{\lambda = (\lambda_0, \ldots, \lambda_m) : \lambda_i \geq 0,\ i = 0, \ldots, k;$$

$$\lambda_i f_i(x) = 0,\ i = 1, \ldots, k;\ \sum \lambda_i \nabla f_i(x) = 0\Big\}.$$

the set of δ-multipliers:

$$\Lambda_\delta(x) = \Big\{\lambda = (\lambda_0, \dots, \lambda_m) \colon \lambda_i \geq 0,\ i = 0, \dots, k;$$
$$\lambda_i f_i(x) = 0,\ i = 1, \dots, k;\ \|\textstyle\sum \lambda_i \nabla f_i(x)\| \leq \delta\Big\},$$

the subset of *normalized* multipliers and δ-multipliers:

$$\Lambda^N(x) = \Big\{\lambda \in \Lambda(x);\ \textstyle\sum |\lambda_i| \leq 1\Big\},$$

$$\Lambda_\delta^N(x) = \Big\{\lambda \in \Lambda_\delta(x);\ \textstyle\sum |\lambda_i| \leq 1\Big\},$$

and the critical cone for (P) at x:

$$K(x) = \{h \colon \nabla f_i(x) h \leq 0,\ i = 0, \dots, k,\ \nabla f_i(x) h = 0,\ i = k+1, \dots, m\}.$$

Now let us say that

(QGC_P) Problem (P) satisfies the quadratic growth condition on S if $f(x)$ defined by (5.20) satisfies (QGC) on S;

(GSO_P) Problem (P) satisfies the general second-order sufficient condition on S if there are a neighbourhood U of S and, regular projection $\pi \colon U \to S$ and an $\alpha > 0$, such that (2.3) is valid with $\Omega_\delta(\pi(x))$ replaced by $\Lambda_\delta^N(\pi(x))$.

(TC_P) For any $x \in S$ and any $i \in I_p(x) := \{i = 1, \dots, k \colon f_i(x) = 0\}$ either $i \in I(x)$ for all $y \in S$ sufficiently close to x, or S and $\{y \colon f_i(y) = 0\}$ are nontangent at x.

Then the theorems are reformulated as follows.

THEOREM 1(P). *The following implications hold*:

$$(GSO_P) \Rightarrow (QGC_P),$$

$$(QGC_P)\ \&\ (TC_P) \Rightarrow (GSO_P).$$

THEOREM 2(P). *Assume that*

(i) *S is a nice compact set of stationary points of* (P) *and f is constant on S*;

(ii) *(P) satisfies (TC_P) on every component of S*,

(iii) *For any $x \in S$ and any $h \in K(x) \setminus T_C(x)$*,

$$\liminf_{u \xrightarrow{S} x} \max_{\lambda \in \Lambda(u)} \mathscr{L}_{xx}(\lambda, x)(h, h) > 0.$$

Then (GSO_P) holds.

THEOREM 3(P). *If (QGC_P) holds, then*

$$\max_{\lambda \in \Lambda^N(x)} \mathscr{L}_{xx}(\lambda, x)(h, h) \geq \beta \operatorname{dist}^2(T_S(x), h), \qquad \forall h \in K(x)\ \forall x \in S,$$

β being the same as on the (QGC_P), in particular,

$$\max_{\lambda \in \Lambda^N(x)} \mathscr{L}_{xx}(\lambda, x)(h, h), \qquad \forall x \in S, \forall h \in K(x) \setminus T_S(x).$$

The corresponding replacement can be also made in all other results.

5.2. Constraint qualification. Further specification of definition and results can be obtained under the assumption that the Mangasarian-Fromovitz constraint qualification holds at any $x \in S$. As S is compact, it follows that there is a constant $\eta > 0$ such that the distance from the origin to the affine manifold spanned by the gradients of the equality constraint functions is greater than η and there is an h in X with the unit norm such that:

$$\nabla f_i(x)h = 0, \qquad i = k+1, \dots, m; \qquad \nabla f_i(x)h \le -\eta, \qquad i \in I_p(x),$$

and

$$\inf\left\{\lambda_0 \colon \lambda \in \Lambda^N(x);\ \sum |\lambda_i| = 1\right\} \ge \eta,$$

which means that the standardly normalized sets of Lagrange multipliers:

$$\Lambda^1(x) = \{\lambda \in \Lambda(x), \lambda_0 = 1\}$$

are uniformly bounded on S. This immediately implies

PROPOSITION 5. *If the (MF) constraint qualification condition is satisfied for any $x \in S$ then in (GSO_P) we can replace $\Lambda^N(x)$ by $\Lambda^1(x)$.*

The change which occurs with the growth condition is more substantial.

PROPOSITION 6. *If the (MF) constant qualification condition is satisfied for all $x \in S$ then (QGC_P) is equivalent to the following:*
(QGC_{MF}) there are a $\beta > 0$ and a neighborhood U of S such that

$$f_0(x) \ge c + \beta \operatorname{dist}^2(S, x),$$

for all feasible $x \in U$.

PROOF. It is clear that $(\mathrm{QGC_P}) \Rightarrow (\mathrm{QGC_{MF}})$. Conversely, assume that $(\mathrm{QGC_{MF}})$ holds. Let

$$A = \{x \colon f_0(x) \le 0, i = 1, \dots, k, f_i(x) = 0, i = k+1, \dots, m\}$$

be the set of feasible elements. It follows from the Robinson regularity theorem (Robinson (1976)) that for any $x \in S$, there are a $\gamma(x) > 0$ and an $L(x) > 0$ such that

$$\operatorname{dist}(A, u) \le L(x) \cdot \max\{f_1(u), \dots, f_0(u), |f_{k+1}(u)|, \dots, |f_m(u)|\}, \tag{5.21}$$

if $\|u - x\| \le \gamma(x)$. As S is compact, we can choose $\gamma > 0$ and $L > 0$ such that (5.2) is valid with $\gamma(x)$ replaced respectively by γ and L. Assuming that $(\mathrm{QGC_P})$ does not hold we shall find a sequence of $\{u^n\}$ outside of S such that $\operatorname{dist}(S, u^n) \to 0$ and

$$f(u^n) \le \frac{1}{n} \operatorname{dist}^2(S, u^n), \tag{5.22}$$

where f is given by (5.20). Then (5.21) implies that

$$\operatorname{dist}(A, u^n) \le \frac{L}{n} \operatorname{dist}^2(S, u^n),$$

hence there is $x^n \in A$ with $\|x^n - u^n\| \leq (L/n)\text{dist}^2(S, u^n)$. Such an x^n cannot belong to S and, in fact, $\text{dist}(S, x^n) \sim \text{dist}(S, u^n)$.

On the other hand, as all functions are Lipschitz continuous near S we have by (5.22)

$$f(x^n) = o(\text{dist}^2(S, u^n)) = o(\text{dist}^2(S, x^n)).$$

Since $x^n \in A \setminus S$, we have

$$\beta\,\text{dist}^2(S, x^n) \leq f_0(x^n) - c \leq o(\text{dist}^2(S, x^n))$$

a contradiction. □

Acknowledgements. The authors are thankful to the referees for helpful remarks. The second author was supported by the Fund of the Promotion of Science at the Technion under Grant 100-820.

References

Ben-Tel, A., J. Zowe (1982). A unified theory of first and second order conditions for extremum problems in topological vector spaces. *Math. Programming Study* **19** 39–76.

Bonnans, J. F. (1992). Directional derivatives of optimal solutions in nonlinear programming. *J. Optim. Theory Appl.* 73.

______, A. D. Ioffe (1995). Quadratic growth and stability in convex programming problems with multiple solutions. *J. Convex Analysis* **2** 41–57.

______, ______, A. Shapiro. Expansion of exact and approximate solutions in nonlinear programming. W. Oettli and P. Pallaschke eds., *Lecture Notes in Economics and Mathematical Systems* 382, Springer Verlag, 103–117.

______, ______, ______ (1992). Développement de solutions exactes et approchées en programmation non linéaire. *Comptes Rendus Acad. Sci. Paris Ser. I*, 119–123.

Gauvin, J., R. Janin (1988). Directional behaviour of optimal solutions in nonlinear mathematical programming. *Math. Oper. Res.* **13** 629–649.

Ioffe, A. D. (1979a). Regular points of Lipschitz functions. *Trans. Amer. Math. Soc.* 61–70.

______ (1979b). Necessary and sufficient conditions for a local minimum. *SIAM J. Control Optim.* **17**, 245–288.

______ (1991). Variational analysis of a composite function: A formula for the lower second order epi-derivative, *Math. Anal. Appl.* **160** 379–405.

______ (1994). On sensitivity analysis of nonlinear program in Banach spaces: The approach via composite unconstrained optimization. *SIAM J. Optim.* **4** 1–43.

Levitin, E. S., A. A. Miljutin, N. P. Osmolowski (1974). On conditions for a local minimum in problems with constraints. *Mathematical Economics and Functional Analysis* B. S. Mitiagin, ed., Nauka, Moscow, 139–202 (In Russian).

Robinson, S. M. (1976). Stability theory for systems of inequalities, Part 2. *SIAM J. Numer. Anal.* **13** 487–513.

Shapiro, A. (1988). Perturbation theory of nonlinear programs when the set of solution is not a singleton. *Appl. Math. Optim.* **18**, 215–229.

______ (1992). Perturbation analysis of optimization problems in Banach spaces. *Numer. Funct. Anal. Optim.* **13**, 97–116.

J. F. Bonnans: INRIA, Domaine de Voluceau, BP 105, 78153 Rocquencourt, France

A. Ioffe: Department of Mathematics, Technion Israel Institute of Technology, Haifa 32000, Israel; e-mail:ioffe@techunix.technion.ac.il