Second-order sufficiency and quadratic growth for non isolated minima

Joseph Frédéric Bonnans
INRIA, Domaine de Voluceau,
BP 105, 78153 Rocquencourt, France.

Alexander Ioffe*
Department of Mathematics
Technion Israel Institute of Technology
Haifa 3200 -Israel

June 23, 2003

Abstract. 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 case when the set of minima may have non isolated points.

Keywords Optimality conditions, Lagrangian, composite functions, transversality, proximal normals, critical cone.

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 [2], [3], [4], [5], [10], [6] 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 [13]

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

*Supported by the Fund of the Promotion of Science at the Technion under the grant 100-820.

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. [1], [8] [11]) is equivalent to (1.1) provided the Mangasarian-Fromovitz constrained qualification is valid [3]. But very little has been known so far about sufficient conditions and (1.1) like estimates in situations when the set of solution has a more complicated structure than just a finite collection of isolated points.

This paper is an attempt to fill the gap. We establish several sufficient condition, 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 the first glance. It requires information which is not "intrinsic" in the sense that it cannot be expressed on terms of derivatives of the Lagrangian function and 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 introduce this condition as the basic sufficiency statement is that it is equivalent to the general growth condition under an additional "transversality" assumption which has a simple and natural formulation.

Transversality considerations 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 "bottleneck" at which all the main difficulties caused by non-unicity 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 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 [8] and sensitivity analysis [10]) is that it allows to get rid of feasibility problems in the course of main arguments.

2 Main results

So 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,

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

the **Lagrangian** of f,

$$\mathcal{S}^n := \{\lambda \in \mathbb{R}^n \; ; \; \lambda \geq 0, \; \sum_{i=1}^{m} \lambda_i : 1\}$$

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

$$\Omega(x) := \{\lambda \in S^m \; ; \; \lambda_i \geq 0, \lambda_i = 0 \text{ if } i \notin I(x); \; \sum \lambda_i \nabla f_i(x) = 0\}$$

(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

$$\Omega_\delta(x) := \{\lambda \in S^m, \; \lambda_i = 0 \text{ if } i \notin I(x); \; \|\sum \lambda_i \nabla f_i(x)\| \leq \delta\}$$

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 $f(x) \equiv$ const $= c_0$ on S.

Definition 1 A mapping π from a neighborhood U of S onto S will be called a *regular projection* onto S if $\pi(x) = x$ for all $x \in S$ and there exists $\varepsilon > 0$ such that

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

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 *transversal* 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 transversal 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 there exists $\beta > 0$ and a neighborhood U of S such that

$$f(x) \geq c_0 + \beta \ \mathrm{dist}^2(S, x) \ , . \forall x \in U \tag{2.2}$$

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

for any $\delta > 0$ there are a neighborhood U of S and, a regular projection $\pi : U \to S$ and an $\alpha > 0$ such that for all $x \in U \backslash S$,

$$\max_{\lambda \in \Omega_\delta(\pi(x))} [\mathcal{L}_x(\lambda, \pi(x))h + \frac{1}{2}\mathcal{L}_{xx}(\lambda, \pi(x))(h, h)] \geq \alpha \|h\|^2. \tag{2.3}$$

where $h = x - \pi(x)$. We also say that f satisfies the **transversality 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 transversal at x.

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

Theorem 1 *The following implications hold :*

$$(GSO) \Rightarrow (QGC),$$

$$(QGC) \& \{(TC) \ on \ S\} \Rightarrow (GSO).$$

Theorem 2 *Suppose that*

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

(ii) *f satisfies (TC) on every component of S,*

(iii) *for any $x \in S$ and any $h \in C(x) \backslash T_S(x)$*

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

Then (GSO) holds.

Theorem 3 *If (QGC) holds, then*

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

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

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

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" [10]. 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 exists $\alpha, \beta > 0$, a neighborhood U of S and a regular projection $\pi : U \to S$ such that for $h := x - \pi(x)$, we have

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

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 $\Omega(y) \subset \Omega_\delta(y)$,

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

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 S4.

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 2are 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*

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

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 $\varepsilon > 0$ such that (2.4) is valid for all $h \in C(x) \cap (T_S^\varepsilon(x) \backslash T_S(x))$, where $T\varepsilon_S(x)$ is the set of approximate contingent directions to S at x:

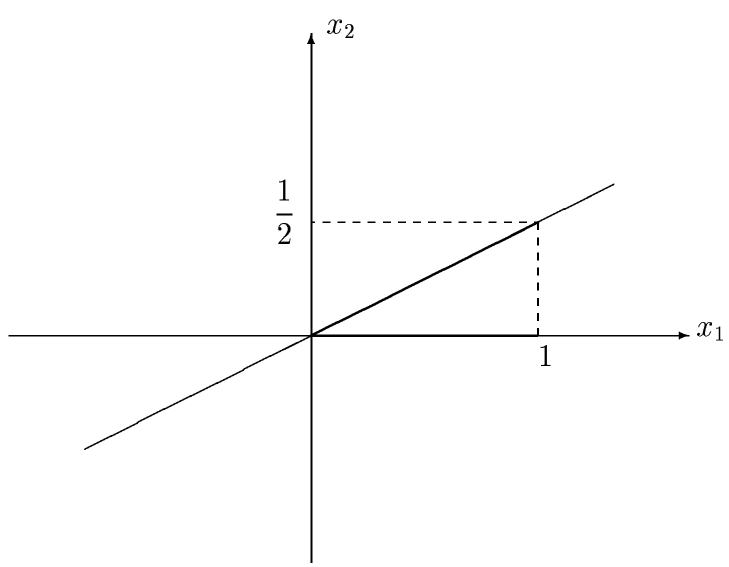
$$T_S^\varepsilon(x) = \{h : \mathrm{dist}(T_S(x), h) \leq \varepsilon \|h\|\}$$

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 picture)

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

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



It can be easily verified that f satisfies (QGC) on S. Indeed :

if $x_2 \leq 0$, $0 \leq x_1 \leq 1$, then $f(x) \geq x_2^2 = \mathrm{dist}^2(S, x)$;

if $x_1 \leq 0$ then

$$\begin{aligned} f(x) &\geq \max\{-x_1, x_2^2 - x_2 x_1\} \geq \max\{x_1^2, |x_2|(|x_2| - |x_1|)\} \\ &\geq \frac{1}{2}[x_1^2 + (x_2^2 - |x_1 x_2|)] \geq \frac{1}{4}(x_1^2 + x_2^2) \geq \frac{1}{4}\|x\|^2; \end{aligned}$$

if $0 \leq x_2 \leq x_1/2$, $0 \leq x_1 \leq 1$, then

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

etc. ...

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

$$\begin{array}{rcl} T_S(0) & = & \{h = (\alpha,\beta) : \alpha \geq 0, \quad \beta = 0 \text{ or } \beta = \alpha/2\} \\ C(0) & = & \{h = (\alpha,\beta) : \alpha \geq 0, \quad \beta \leq \alpha/2\}. \end{array}$$

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

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

Now taking $h = (\alpha,\beta) \in C(0)\backslash T_S(0)$ which means that $\beta \neq 0$ (and $\beta < \alpha/2$) we see that

$$\liminf_{\substack{x \to 0 \\ S}} \max_{\lambda\in\Omega(x)} \mathcal{L}_{xx}(\lambda,x)(h,h) \leq -\frac{1}{2}\beta^2 < 0.$$

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

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

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 contigent 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 $\mathrm{dist}(x(t), S) = t$, 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 $\lim\limits_{\substack{S \\ u \to x}} \Omega(u)$ 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, $\Omega(.)$ is no more then 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 $i \in I(x) \backslash I_0(x)$ such that $\nabla f_i(x) \neq 0$, we have $\lambda_i = 0$. In other words, "no new multipiers are added by new constraints".

Proof. Pick $i \in I(x) \backslash I_0(x)$. 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

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

On the other hand, if $\nabla f_i(x) \neq 0$, then the equality $\nabla f_i(x) h = 0$ (meaning that h is on 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 $0 = \sum_{i=1}^{m} \lambda_i \nabla f_i(x).d$ 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, and set for $u \in S$*

$$B(u) := \{v : I(v) \backslash I(u) \neq \emptyset\}.$$

Let $x \in D$. Then there exists $\gamma > 0$ such that whenever $x^n \in D$ and $x^n \to x$, then

$$\mathrm{dist}(B(x^n), x^n) \geq \gamma \|x - x^n\|.$$

Proof. Assuming the contrary, we find a sequence of $x^n \in D$ converging to an $x \in D$ and such that

$$\mathrm{dist}(B(x^n), x^n) = o(\|x - x^n\|).$$

Extracting if necessary a subsequence, we may assume that $I(x^k)$ is a constant set say J, and that there exists $i \in I(x) \backslash J$ and a sequence $y^n \to x$ such that $I(y^n) \ni i$ and $\|y^n - x^n\| = O(\|x - x^n\|)$. It follows that the limit of $(x^n - x)/\|x^n - x\|$ (which we may assume to exist, extracting a further subsequence) is equal to the limit of $(y^n - x)/\|y^n - x\|$. This common limit is in $T_D(x)$ as well as in the contingent cone to $\{y \ ; \ f_i(y) = f_i(x)\}$, and of norm 1, in contradiction with (TC). □

Proof. of Theorem 1

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

$$|\mathcal{L}(\lambda, x+h) - \mathcal{L}(\lambda, x) - \mathcal{L}_x(\lambda, x)h - \frac{1}{2}\mathcal{L}_{xx}(\lambda, x)(h,h)| < (\alpha - \beta)\|h\|^2$$

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

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

i.e. (QGC) holds.

(QGC) & {(TC) on S} ⇒ (GSO) . Assume the contrary. Then as S is compact there exists a $\delta > 0$ and $x^n \xrightarrow{S} x$ such that if a sequence $u \in S$ satisfies $\|u^m - x^n\| \leq n\ \mathrm{dist}(S, x^n)$ we have

$$\max_{\lambda \in \Omega_\delta(u)} (\mathcal{L}_x(\lambda, u)h + \frac{1}{2}\mathcal{L}_{xx}(\lambda, u)(h,h)) \leq \frac{1}{n}\|h\|^2 \tag{3.6}$$

where $h^n = x^n - u$. We fix u^n ins the following way
(A) $\mathrm{dist}(S, x^n) = o(\|x - x^n\|)$. Let $u^n \in S$ be such that $\|x^n - u^n\| = \mathrm{dist}(S, x^n)$; set $h^n = x^n - u^n$. By Lemma 1, $I(x^n) \subset I(u^n)$ for large n.

(B) There exists a $\theta > 0$ such that, extracting if necessary a subsequence we may assume that

$$\mathrm{dist}(S, x^n) \geq \theta\|x^n - x\|.$$

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

$$\begin{aligned} \beta\|h^n\|^2 &\leq f(u^n + h^n) - f(u^n) \\ &= \max_{i \in I(u^n)} \{f_i(u^n + h^n) - f_i(u^n)\} \\ &= \max_{\lambda \in \Omega_\infty(u^n)} \{\sum \lambda_i (f_i(x^n + h^n) - f_i(u^n))\} \\ &= \max_{\lambda \in \Omega_\infty(u^n)} \{\mathcal{L}_x(\lambda, u^n)h^n + \frac{1}{2}\mathcal{L}_{xx}(\lambda, u^n)(h^n, h^n)\} + o(\|h^n\|^2). \end{aligned} \tag{3.7}$$

Set

$$\xi := \max\{\|\mathcal{L}_x(\lambda, x)\|; x \in S, \quad \lambda \in \Omega_\infty(x)\}.$$

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

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

so with (3.6) and (3.7)

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

which may only happen if $\beta = 0$ contrary to (QGC). □

We now prove Theorem 2. The proof is based on the following lemma.
Proof.

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

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

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 $\nabla f_i(x)e \le 0$ for all $i \in I(x)\backslash J$. Moreover, given $\varepsilon > 0$, the sequence of w^n can be chosen in such a way that $\|h - e\| < \varepsilon$, 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 $\|u - x^n\| \le n\,\mathrm{dist}(S, x^n)$, (3.6) holds.

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

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

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

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

$$\max_{\lambda\in\Omega_\delta(x)}\mathcal{L}_x(\lambda,x)h\le 0.$$

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

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

On the other hand it follows from (3.9) that

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

and therefore

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

Hence we arrived at a contradiction.

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

Assume for the moment that the lemma has been already 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 $\operatorname{dist}(S, x^n)$, we have $\|x^n - w^n\| \leq n \operatorname{dist}(S, x^n)$ so that (3.6) holds with $x = w^n$ and $h = x^n - w^n$, that is to say

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

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

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

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

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

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

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

Let y^n and v^n be defined by

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

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

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

and the sets $S_j = Q(C_j)$ are transversal 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 $i \in I(x)\backslash J$ 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 $i \in I(x)\backslash J$, $\nabla f_i(x) \ne 0$. We can assume that $\gamma < \varepsilon$. Fix $M > 1 + 2\gamma^{-1}$ and let

$$\begin{aligned} \alpha^n &:= M\|x^n - u^n\|/\|x^n - x\|, \\ z^n &:= (1-\alpha^n)v^n, \\ w^n &:= Q(z^n). \end{aligned} \tag{3.12}$$

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 $\|h - e\| \le \varepsilon$. 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

$$\begin{aligned} w^n &= \alpha^n x + (1-\alpha^n)u^n + [Q'(0) - Q'(v^n)]\alpha^n(1-\alpha^n)v^n + o(\alpha^n\|v^n\|), \\ &= \alpha^n x + (1-\alpha^n)u^n + o(\alpha^n\|v^n\|), \\ &= \alpha^n x + (1-\alpha^n)u^n + o(\|u^n - x^n\|), \end{aligned}$$

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, $\|w^n - x^n\| = O(\text{dist}(S_1, x^n))$ 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 $g^n = \dfrac{x^n - u^n}{\|x^n - u^n\|}$:

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

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) :

$$\|e - h\| \leq \frac{2}{M-1} < \gamma.$$

Q.E.D.

We end this section by proving Theorem 3.

Proof. We have (see e.g. [9], Corollary 5)

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

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

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

hence

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

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 $PN(S, x)$ the collection of proximal normals to S at x. It is always a closed convex cone.

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

$$C_\varepsilon(x) = \{h : \nabla f_i(x)h \leq \varepsilon\|h\|, \quad i \in I(x)\}.$$

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

$$\max_{\lambda \in \Omega_\delta(x^n)} [\mathcal{L}_x(\lambda, x^n)h^n + \frac{1}{2}\mathcal{L}_{xx}(\lambda, x^n)(h^n, h^n)] \leq O(\|h^n\|^2).$$

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

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

$$\max_{\lambda \in \Omega_\infty(x^n)} [\mathcal{L}_x(\lambda, x^n)h^n + \frac{1}{2}\mathcal{L}_{xx}(\lambda, x^n)(h^n, h^n)] \leq O(\|h^n\|^2),$$

hence

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

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

Proposition 2 *Suppose that there exists $\varepsilon > 0$ and $\alpha > 0$ such that*

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

for any $x \in S$ and any $h \in C_\varepsilon(x) \cap PN(S, x)$. 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, $h^n \in C_\varepsilon(x^n)$ if n is large enough and, by definition $h^n \in PN(S, x^n)$. 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

$$\begin{aligned}
I_1(x; h) &= \{i \in I(x), \nabla f_i(x)h = f'(x; h)\};\\
M(x; h) &= \{\lambda :\in S^m;\ \lambda_i = 0\ if\ i \notin I_1(x; h);\ \mathcal{L}_x(\lambda, x)h \ge 0\};\\
\mu(x; h) &= \min\{\|\mathcal{L}_x(\lambda, x)\| : \lambda \in M(x; h)\}.
\end{aligned}$$

We also set

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

Proposition 3 *Suppose that there exists $\bar\mu > 0$ such that*

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

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

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

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 $C_\varepsilon(x) \cap PN(S,x)$. It follows from [7] that

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

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

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

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

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

if $x \in S$, $\lambda_i \geq 0$, $\sum \lambda_i = 1$, $\|h\| = 1$, $\|h - h'\| \leq \delta_1$. Let $\varepsilon > 0$ be so small that

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

By (4.16) and (4.18) for any $h \in C_\varepsilon(x) \cap PN(S,x)$ there is a $e \in C(x)$ such that

$$\|h - e\| \leq \frac{\varepsilon}{\bar{\mu}} \|h\| \leq \delta \|e\|. \tag{4.19}$$

This means that $e \in C(x) \cap PN_\delta(S,x)$. Then by hypothesis

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

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

$$\begin{aligned} \max_{\lambda \in \Omega(x)} \mathcal{L}_{xx}(\lambda, x)(h,h) &\geq \max_{\lambda \in \Omega(x)} \mathcal{L}_{xx}(\lambda, x)(e,e) - \frac{\alpha}{2}\|h\|^2 \\ &\geq \alpha(|e|^2 - 1/2)\|h\|^2). \end{aligned}$$

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

It can be be observed that Propositions 1 and 2, though much simpler to formulate, are weaker results then 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

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

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 non-linear programs:

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

The very fact that theorems on a maximum function as considered above can be applied to (P) follows from the simple observation (cf. [10]) :

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

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

Then the following two properties are equivalent :

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

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

Proof. The implication (b) $\Rightarrow$ (a) is obvious. Conversely, if (a) holds, then $f(x) \ge f_0(x) > 0$ for any feasible $x \in U \backslash S$. On the other hand if x is not feasible then either $f(x) > 0$ for some $i = 1, \dots, k$, or $|f_i(x)| > 0$ for some $i = k+1, \dots, m$; 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 on (a) replacing $|f_i(x)|$ by $\max\{f_i(x), -f_i(x)\}$ on (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 $i = k+1, \dots, m$ are constant on S.

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

$$\Lambda(x) = \{\lambda = (\lambda_0, \dots, \lambda_m) : \lambda_i \ge 0,\ i = 0, \dots, k;\ \lambda_i f_i(x) = 0,\ i = 1, \dots, k;\ \sum \lambda_i \nabla f_i(x) = 0\},$$

the set of δ-multipliers:

$$\Lambda_\delta(x) = \{\lambda = (\lambda_0, \dots, \lambda_m) : \lambda_i \ge 0,\ i = 0, \dots, k,\ \lambda_i f_i(x) = 0,\ i = 1, \dots, k;\ \|\sum \lambda_i \nabla f_i(x)\| \le \delta\},$$

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

$$\begin{aligned} \Lambda^N(x) &= \{\lambda \in \Lambda(x)\ ;\ \sum |\lambda_i| \leq 1\} \\ \Lambda^N_\delta(x) &= \{\lambda \in \Lambda_\delta(x)\ ;\ \sum |\lambda_i| \leq 1\}. \end{aligned}$$

and the critical cone for (P) at x:

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

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 neighborhood U of S and, regular projection $\pi : U \to S$ and an $\alpha > 0$, such that (2.3) is valid with $\Omega_\delta(\pi(x))$ replaced by $\Lambda^N_\delta(\pi(x))$.

(TC_P) For any $x \in S$ and any $i \in I_P(x) := \{i = 1, \ldots, k : f_i(x) = 0\}$ either $i \in I(x)$ for all $y \in S$ sufficiently close to x, or S and $\{y : f_i(y) = 0\}$ are transversal 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 $h \in K(x) \backslash T_C(x)$*

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

Then (GSO_P) holds.

Theorem 3 (P) *If (QGC_p) holds, then*

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

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

$$\max_{\lambda \in \Lambda^N(x)} \mathcal{L}_{xx}(\lambda, x)(h, h), \quad \forall x \in S, \quad \forall h \in K(x) \backslash T_S(x).$$

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 :

$$\nabla f_i(x)h = 0, \quad i = k+1, \ldots, m \; ; \; \nabla f_i(x)h \leq -\eta, \quad i \in I_p(x),$$

and

$$\inf\{\lambda_0 : \lambda \in \Lambda^N(x) \; ; \; \sum |\lambda_i| = 1\} \geq \eta$$

which means that the standardly normalized sets of Lagrange multipliers:

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

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

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

for all feasible $x \in U$.

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

$$A = \{x : f_0(x) \leq 0, \quad i = 1, \ldots, k, \quad f_i(x) = 0, \quad i = k+1, \ldots, m\}$$

be the set of feasible elements. It follows from the Robinson regularity theorem [12] that for any $x \in S$, there are a $\gamma(x) > 0$ and an $L(x) > 0$ such that

$$\text{dist}(A, u) \leq L(x). \max\{f_1(u), \ldots f_0(u), \; |f_{k+1}(u)|, \ldots, |f_m(u)|\} \tag{5.21}$$

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

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

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

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

hence there is $x^n \in A$ with $\|x^n - u^n\| \leq \frac{L}{n} \text{dist}^2(S, u^n)$. 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 $x^n \in A \backslash S$, 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. □

References

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

[2] J.F. Bonnans : A review of some recent results in the perturbation theory of nonlinear programs. Actas del XII C.E.D.Y.A., II Congreso de Matematica Aplicada, Universidad de Oviedo, Sept. 23-27, 1991, 3-19.

[3] J.F. Bonnans : Directional derivatives of optimal solutions in nonlinear programming, JOTA 73 (1992).

[4] J.F. Bonnans, A.D. Ioffe, A. Shapiro : Expansion of exact and approximate solutions in nonlinear programming. In Lecture Notes in Economics and Mathematical Systems 382, W. Oettli and P. Pallaschke eds., Springer Verlag, 103-117.

[5] J.F. Bonnans, A.D. Ioffe, A. Shapiro (1992) : Développement de solutions exactes et approchées en programmation non linéaire. Comptes Rendus Acad. Sci. Paris Ser. I, 119-123.

[6] J. Gauvin, R. Janin : Directional behaviour of optimal solutions in nonlinear mathematical programming. Math. Oper. Res. 13(1988), 629-649.

[7] A.D. Ioffe : Regular points of Lipschitz functions, Trans. Amer. Math. Soc. (1979), 61-70.

[8] A.D. Ioffe (1979) : Necessary and sufficient conditions for a local minimum. SIAM J. Control Optimiz. 17, 245-288.

[9] A.D. Ioffe : Variational analysis of a composite function : a formula for the lower second order epi-derivative, Math. Anal. Appl. 160(1991), 379-405.

[10] A.D. Ioffe : On sensitivity analysis of nonlinear program in Banach spaces : the approach via composite unconstrained optimization, SIAM J. Optimization, to appear.

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

[12] S.M. Robinson : Stability theory for systems of inequalities, part 2, SIAM J. Numer. Analysis, 13 (1976), 487-513.

[13] A. Shapiro (1988) : Perturbation theory of nonlinear programs when the set of solution is not a singleton. Appl. Math. Optim. 18, 215-229.