A Semistrong Sufficiency Condition for Optimality in Nonconvex Programming and Its Connection to the Perturbation Problem

J. F. Manso

Communicated by A. V. Fiacco

Abstract: What happens when a nonconvex program has a local solution that is not a global minimizer? Under the relatively weak second-order assumption (differentiabil- ity and bounded level sets, and no inflection points), the minimizing optimal set can be obtained as a perturbation of the minimizing optimal set of the convex relaxed problem. A sufficient condition is given for the minimizing optimal set to be the global minimizer of the original problem.

1. Introduction

Let us consider for every $c \in R^m$, the following nonlinear program:

$$\min \{ f(x) : x \in X \}$$

subject to

$$g_i(x) = 0, \quad i = 1, \ldots, m$$

$$h_j(x) \leq 0, \quad j = 1, \ldots, p$$

where $f, g_1, \ldots, g_m, h_1, \ldots, h_p : R^n \to R$ are continuously differentiable functions, and $X = \{ x \in R^n : h_1(x) \leq 0, \ldots, h_p(x) \leq 0 \}$.

We denote $F(c) = \{ x \in X : f(x) = \inf \{ f(z) : z \in X \} \}$, the set of minimizers for the problem, and $C(c)$, the set of critical points of $f$ with respect to $c$. We define $C(c) = \{ x \in X : \nabla f(x) \cdot \nabla g(x) = 0 \}$.

The following holds for the sufficient and necessary conditions:

- If $c \in C(c)$, then $f(x) \geq f(c)$ for all $x \in X$.
- If $c \notin C(c)$, then $f(x) > f(c)$ for some $x \in X$.

This paper extends the results of Manso and Fiacco (1977) to the case of nonconvex programs.
Let $x^*$ be a local solution of (1)$_n$, and suppose that
\[ \nabla g_i(x^*_n) = 0, \quad i = 1, \ldots, p. \]
are linearly independent. (2)

Then there exists a unique Lagrange multiplier $\lambda^*$ in $R^p$ such that
\[ (i) \quad \nabla f(x^*)_n + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*_n) = 0, \]
\[ (ii) \quad g_i(x^*)_n = 0, \quad i = 1, \ldots, m, \]
\[ (iii) \quad g_i(x^*) < 0, \quad \lambda_i^* > 0, \quad \lambda_i^* g_i(x^*)_n = 0, \quad i = m + 1, \ldots, p. \]

As we are interested in what happens in the neighborhood of $x^*$, we may suppose, to simplify the study, that
\[ g_i(x^*_n) = 0, \quad i = 1, \ldots, p. \]

Let us denote
\[ \mu^*_n = (i = m + 1, \ldots, p; \lambda_i^* = 0), \]
\[ H = \nabla f(x^*)_n + \sum_{i=1}^{p} \lambda_i^* \nabla g_i(x^*_n). \]

A necessary condition for $x^*$ to be a local solution of (1)$_n$, called the second-order necessary condition, is that
\[ \nabla g_i(x^*)_n d = 0, \quad i = (1, p); \mu^*_n \]
\[ \nabla g_i(x^*)_n d = 0, \quad i = \mu^*_n \]
implies $d' Hd > 0$. (4c)

By changing an inequality into a strict inequality, we obtain the standard second-order sufficiency condition:
\[ d \neq 0, \]
\[ \nabla g_i(x^*)_n d = 0, \quad i = (1, p); \mu^*_n \]
\[ \nabla g_i(x^*)_n d = 0, \quad i = \mu^*_n \]
implies $d' Hd > 0$. (5d)

Condition (5) ensures that $x^*$ is a strict local solution of (1)$_n$. Together with (2), it implies the following stability result (see Ref. 1, Theorem 3.1, and Corollary 4.3).

**Theorem 1.1.** Let $x^*$ be a local solution of (1)$_n$, and let $x^*$ be an associated Lagrange multiplier such that (2), (3), and (5) hold. Then there exists $\epsilon > 0$ and $\alpha > 0$ such that, for $|x - x^*| < \epsilon$, problem (1)$_n$ possesses (at least) one local solution $x^*$, with an associated Lagrange multiplier $\lambda^*$, such that
\[ ||x - x^*|| + ||\lambda - \lambda^*|| < \alpha. \]

In addition, there exists $\alpha > 0$ such that, for $|x - x^*| < \epsilon$, any local solution $x^*$ such that $||x - x^*|| < \alpha$ has an associated Lagrange multiplier $\lambda^*$ that satisfies (6).

Stronger results have been obtained with the hypothesis that the following strong second-order sufficiency condition holds:
\[ d \neq 0, \]
\[ \nabla g_i(x^*)_n d = 0, \quad i = (1, p); \mu^*_n \]
implies $d' Hd > 0$. (7c)

We notice that (7) implies (5). Using (7), Jittorntru (Ref. 2) proved the following theorem.

**Theorem 1.2.** Assume that the hypotheses of Theorem 1.1 hold, with (7) instead of (5). Then for $x$ small enough, there is a unique solution $x^*$ of (1)$_n$, such that (6) holds, and the mapping $x \rightarrow (x^*, \lambda^*)$ has directional derivatives.

If $\mu^*_n$ is empty, i.e., the strict complementarity conditions are fulfilled, (5) and (7) are equivalent and the mapping $x \rightarrow (x^*, \lambda^*)$ is $C^1$ (see Ref. 3). It is easy to see that, if card ($\mu^*_n$) = 1, conditions (5) and (7) are also equivalent; but this is no longer true, in general, if card ($\mu^*_n$) > 2. The main drawback of the strong second-order sufficiency condition is that it assumes the strict positivity of an amount that can be strictly negative for an isolated minimum. Hence, it is very strong. In contrast, we will use an hypothesis in which we only assume the strict positivity of nonnegative amounts.

**Definition 1.1.** We will say that $L \subset (1, p)$ is a pseudo-optimal subpartition of (1), or shortly a pseudo-optimal subpartition, if
\[ I_1 \supset (1, p) \setminus L \]
\[ L \subset (1, p) \setminus I_1, \]
and
\[ \nabla g_i(x)^{d} = 0, \quad i \in I_1, \]
\[ \nabla g_i(x)^{d} \neq 0, \quad i \in I_2, \]
imply \( d' Hd > 0 \).

**Definition 1.2.** We say that \( x^* \) satisfies the semistrong second-order sufficiency condition if the following holds: For any pseudo-optimal sub-partition \( (I_1, I_2) \), we have
\[ d = 0, \]
\[ \nabla g_i(x)^{d} = 0, \quad i \in I_1, \]
\[ \nabla g_i(x)^{d} \neq 0, \quad i \in I_2, \]
imply \( d' Hd > 0 \).

Let us remark that \( (I_1, p) - I_2^\infty, I_2^\infty \) is a pseudo-optimal sub-partition. Hence, (8) is stronger than (5). On other hand, the semistrong second-order sufficiency condition is obviously weaker than the strong second-order sufficiency condition. If the problem \((1), k)\) is concave (i.e., if \( f \) and \( g_i, i = 1, \ldots, m \), are linear), then \((I_1, p) - I_2^\infty, I_2^\infty \) is a pseudo-optimal sub-partition. Hence, the strong and semistrong conditions are equivalent for convex problems. We stress the fact that the hypothesis is made on \( f \) and \( g \) but is independent of the perturbations.

The paper is organized as follows. In Section 2, we assume only that the standard second-order sufficiency condition holds at \( x^* \) and, considering a primal-dual local solution \((x^*, \lambda^*)\) of \((1)\), we study the limit points of \( e(x^*, \lambda^*) \). We show that the limit points do not necessarily correspond to the local minima of the tangent quadratic problem (an example where this curious phenomenon occurs is given in Section 4). We call these limit points pseudo-solutions of the tangent quadratic problem.

In Section 3, we study how to compute the local minima associated with a given nonnegative pseudosolution of the tangent quadratic problem, and we deduce the characterization of the finite set of the local minima of \( e(x^*, \lambda^*) \), close to \( x^* \), when the semistrong second-order sufficiency condition holds.

2. Some Results Using Only the Standard Second-Order Sufficiency Condition

As we are interested in the sequel in the directional derivatives of the solutions of \((1)\), we limit our analysis to the case \( d = 0 \). We define a branch to be a mapping \( \epsilon \rightarrow (x^*(\epsilon), \lambda^*(\epsilon)) \), defined for \( \epsilon \) small enough, such that \( x(\epsilon) = (x^*(\epsilon), \lambda^*(\epsilon)) \) when \( \epsilon = 0 \). We will speak of a branch of solutions if \( x^* \) is a local solution of \((1)\), and \( \lambda^* \) is its (uniquely defined for \( \epsilon \) small and \( x^* \) close to \( x^0 \)) associated Lagrange multiplier, i.e.,
\[ \begin{align*}
(i) & \quad \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^*) = 0, \\
(ii) & \quad g_i(x^*) = 0, \quad i = 1, \ldots, m, \\
(iii) & \quad g_i(x^*) \leq 0, \quad \lambda_i \geq 0, \quad \lambda_i g_i(x^*) = 0, \\
& \quad i = m + 1, \ldots, p.
\end{align*}
\]
Set
\[ d^* = e(x^* - x^0), \quad \mu^* = e(x^* - x^0). \]
Under the hypotheses of Theorem 1.1, \((d^*, \mu^*)\) is bounded. Denote
\[ \alpha = d^* e(x^* - x^0) / \delta(x^* - x^0), \quad \alpha = g_i(x^*) / \delta(x^* - x^0), \]
\[ i = 1, \ldots, p. \]
The usual linearization techniques applied to the system \((9)\) lead to the following system (see Ref. 2):
\[ \begin{align*}
Hd + \alpha \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) & = 0, \\
\nabla g_i(x^*) d & = 0, \quad i \in (1, p) - I_2^\infty, \\
\nabla g_i(x^*) d & = \alpha, \quad \mu_i > 0, \quad \mu_i (\nabla g_i(x^*) d) = 0, \\
& \quad i \in I_2^\infty.
\end{align*}
\]
For any solution \((d, \mu)\) of (10), we will denote
\[ \lambda^*_d = (\epsilon (1, p) - I_2^\infty, \mu_d = 0), \]
\[ I_d = \{ i \in (1, p) : \mu_i = 0 \}. \]
Some simple computations involving the difference between (3) and (9) imply the following proposition.

**Proposition 2.1.** We assume that \((2), (3), \) and (5) hold. Let \((d, \mu)\) be a limit point of \((d^*, \mu^*)\) when \( \epsilon \rightarrow 0 \). Then \((d, \mu)\) satisfies (10).

Now, let us analyze what supplementary information on \((d, \mu)\) is given by the second-order necessary condition satisfied by \((x^*, \lambda^*)\). We denote
\[ I_d = \{ i = m + 1, \ldots, p : g_i(x^*) = 0, \lambda_i = 0 \}. \]
The second-order necessary condition for \( x^* \) is

\[
\forall g(x')d = 0, \quad i \in I^*(x') - I(x'),
\]

\[
\forall g(x')d = 0, \quad i \in I_1(x'),
\]

(1a)

(1b)

\[
\text{implies} \quad d' H d \neq 0,
\]

(1c)

where

\[
H = \nabla^2 g(x') + \frac{1}{\mu} \lambda^T \nabla g(x').
\]

Proposition 2.2. Assume that (2), (3), and (5) hold. Let \( (x_k)_{k=0} \) be a sequence such that \( x_k \to 0 \), and let \( (x^*)^1 \) be a sequence of local solutions of (1) with associated unique Lagrange multipliers \( \lambda^* \). Let \( (I_k, I_1) \) be such that

\[
(I_k, I_1) = (I^*(x^*) - I(x^*), I_1^*(x^*) - I_1(x^*)).
\]

Then, \( (I_k, I_1) \) is a pseudo-optimal partition; and, if \( (d, \mu) \) is a limit point of \( (x')^* (x' - x'), \lambda^* - \lambda^* \), then

\[
J(d) = k_1(d) \subseteq I_1 \cap J(d),
\]

(12a)

\[
I_2 \subseteq J(d) - I_1,
\]

(12b)

Proof. Since (2) holds, the condition (11) holds at \( x^* \), and thus

\[
\forall g(x')d = 0, \quad i \in I_1,
\]

\[
\forall g(x')d = 0, \quad i \in I_1,
\]

\[
\text{implies} \quad d' HS > 0.
\]

Let \( (d, \mu) \) be a limit point of

\[
(d^*, \mu^*) = (x')^* (x' - x^*), \lambda^* - \lambda^*.
\]

We have that \( I_k \subseteq J(d^*) \), and \( \lambda^* \neq 0 \), for \( x^* \) small enough, if \( i > m \) and \( i \in J(d) - J(d') \). Hence, \( J(d) - J(d') \subseteq I_1 \). As \( I_1 \subseteq \mathbb{R}^n \), we deduce that

\[
I_1 \subseteq J(d) - I_1 = I_2 - I_1.
\]

It is usual to connect the system (10) to the following tangent quadratic problem (TQP):

\[
\text{min } c(d + d' H d),
\]

(13a)

\[
\forall g(x')d + c = 0, \quad i \in I_1,
\]

(13b)

\[
\forall g(x')d + c = 0, \quad i \in I_2.
\]

(13c)

Problem (13) is obtained from (TQP) by discarding some inequality constraints and by converting some others into equality constraints.

(ii) We remark that a pseudo-solution of the (TQP) may be associated with several pseudo-optimal subparts.

(iii) A pseudo-solution of the (TQP) is a stationary point of (TQP), since it satisfies (10); but the converse is not always true, as shown in Example 4.2.

3. Nondegenerate Pseudo-Solutions of the (TQP) and the Semistrong Second-Order Sufficiency Condition

This section contains the main results. We first present some results using a nondegeneracy hypothesis on a given pseudo-solution of the (TQP). Then, we characterize the set of solutions of the perturbed problem under the semistrong second-order sufficiency condition.
Definition 3.1. A pseudo-optimal subpartition \( (I_1, I_2) \) is nondegenerate if the following condition holds:

\[
\begin{align*}
\delta &\neq 0, \\
\forall \delta \in (x') \delta = 0, &\quad i \in I_1, \\
\forall \delta \in (x') \delta < 0, &\quad i \in I_2, \\
\text{imply} &\quad \delta H \delta > 0.
\end{align*}
\]

If, in addition, \( (I_1, I_2) \) is associated with a pseudo-solution \( d \) of the TQP, we say that \( d \) is a nondegenerate pseudo-solution of the (TQP) and the \( (d, I_1, I_2) \) is nondegenerate.

Remark 3.1. There is at most one pseudo-solution of the (TQP) associated with a nondegenerate pseudo-optimal subpartition \( (I_1, I_2) \) because it is a part of the solution \( (d, \delta) \) of the linear system

\[
Hd + \sum_{i \in I_1} \mu_i \cdot g_i(x') = -c_0,
\]

\[
g_i(x') \delta = c_i, \quad i \in I_1 \cup I_2,
\]

whose solution is unique, since (2) and (14) hold.

We consider the related nonlinear system

\[
\begin{align*}
\nabla f(x') + \sum_{i \in I_1} \lambda_i \nabla g_i(x') &= 0, \\
g'(x') &= 0, \quad i \in I_1 \cup I_2.
\end{align*}
\]

Lemma 3.1. We assume that (2), (3), and (14) hold. Then, for \( \epsilon \) small enough, there is a unique \( C^1 \)-branch \( (x', \lambda') \) such that (15) is satisfied and \( \lambda_i = 0, \quad i \in (1, p) \setminus (I_1 \cup I_2) \).

Proof. The system (15) is satisfied as \( \epsilon = 0 \) by \( (x', \lambda') \). By Remark 3.1, the Jacobian of the system (15) with respect to \( (x, \lambda_i, i \in I_1 \cup I_2) \) is nondegenerate. Hence, by the implicit function theorem, the system (15) has, for \( \epsilon \) small, a unique solution \( (x', \lambda'), i \in I_1 \cup I_2 \). Taking \( \lambda_i = 0, \quad i \in (1, p) \setminus (I_1 \cup I_2) \), we get the result.

We will say that this branch is associated with \( (d, I_1, I_2) \). The following proposition gives a means to recognize whether this branch corresponds to the solutions of \( (I_1, I_2) \).

Proposition 3.1. We assume that (2) and (3) hold. Let \( (x', \lambda') \) be a branch associated with a nondegenerate triple \( (d, I_1, I_2) \). Then, for \( \epsilon \) small enough, necessary conditions for \( x' \) to be a local minimum of (1), are

\[
\begin{align*}
\delta x' &= 0, \\
\lambda_i &\neq 0, \quad i \in I_2 \setminus (I_1 \cup I_2), \\
(\delta x') - I_2 (x') &= \text{a pseudo-optimal subpartition},
\end{align*}
\]

and sufficient conditions are that (16) holds and

\[
(\delta x') - I_2 (x') \text{ is a pseudo-optimal subpartition}.
\]

Proof. The definition of \( (J_i d) \) implies that, for \( \epsilon \) small, \( g_i(x') = 0 \), if \( i \in \overline{d} \); and, by the definition of that branch and (12), \( g_i(x') = 0 \), for \( i \in \overline{J_i d} \) and \( I_2 \). Hence, \( x' \) will be feasible for (1), iff (16a) holds. In the same way, we see that as \( \lambda_i = 0 \), if \( i \notin I_1 \cup I_2 \) and \( \lambda_i > 0 \), if \( i = 1 \), \( i = p > 1 \) and \( i \notin I_2 \), then \( \lambda' \) will be feasible iff \( I_2 \) holds. As \( \lambda' \) is the only possible multiplier associated with \( x' \), condition (16) is necessary. If \( x' \) is a local solution of (1), the necessity of (17) for \( \epsilon \) small enough is a consequence of Proposition 2.20. Let us now prove that (16) and (18) are sufficient conditions. If \( x' \) satisfies (16) but is not a local solution, then for some \( \delta x' \neq 0 \):

\[
\begin{align*}
\forall \delta x' \neq 0, &\quad i \in I_2 (x') - I_2 (x'), \\
\forall \delta x' \neq 0, &\quad i \notin I_2 (x'), \\
(\delta x') - H \delta x' &\neq 0.
\end{align*}
\]

We may suppose that \( \delta x' = 0 \). We take a converging subsequence of each \( \delta x' \), as \( \epsilon \to 0 \). Passing to the limit in the above relations, we deduce that no limit set of \( (\delta x') - I_2 (x'), I_2 (x') \) can be a nondegenerate pseudo-optimal subpartition. This proves that (16) and (18) are sufficient conditions for optimality.

Corollary 3.1. We assume that (2), (3), and the semistable second-order sufficiency hold \( (8) \) hold at \( x' \). Let \( (x', \lambda') \) be as in Proposition 3.1. Then, for \( \epsilon \) small enough and \( x' \) close to \( x', (16) \) and (18) are a necessary and sufficient condition for \( x' \) to be a local solution of (1).
that is not a local solution of the (TQP). We consider the following problem:

\[ \min x_1^2 - x_2^2 + (\epsilon^2 - 2/3) (x_2^2 + x_2) \]
\[ 2x_1 - x_2 - \epsilon = 0 \]
\[ -2x_1 - x_2 - \epsilon = 0 \]

For \( \epsilon = 0 \), the problem reduces to

\[ \min x_1^2 - x_2^2, x_1 \neq 2x_2 \]

It has a unique minimum \( x^* = (0, 0) \) with an associated unique multiplier \( x^* = (0, 0) \).

Let us consider the tangent quadratic problem at \( x^* = 0 \):
\[ \begin{aligned}
2d_1 - d_2 - 1 &= 0, \\
-2d_1 - d_2 - 1 &= 0,
\end{aligned} \]

The first-order optimality conditions for problem (21) are

\[ \begin{aligned}
2d_1 - d_2 - 1 &= 0, \\
-2d_1 - d_2 - 1 &= 0,
\end{aligned} \]

Equations (22) have two solutions: (i) a solution for which only the first constraint is binding, we get
\[ d = (4/3, 0), \mu = (4, 0) \]

and (ii) a solution for which only the second constraint is binding; we get
\[ d = (-1, 0), \mu = (4, 0) \]

The first solution \( d \) is in fact the only local minimum of (21) and there is a branch of solutions of (19), associated with \( d \):
\[ s_1^* = (16/9 - 5e^2/3, 11e/9 - 4e^2/3) \]
\[ x_1^* = (16/9 - 5e^2/3, 0) \]

In contrast, the second solution \( d \) is only a saddle point of problem (21). It is also a pseudo-solution associated to the subpartition \((2), (2)\). There is actually a branch of solutions associated with \( d \):
\[ s_1^* = (-2e/3, 0) \]
\[ x_1^* = (0, e) \]
Example 4.2. In this example, the (TQP) has a stationary point \( d \) is not a pseudo-solution. The problem is

\[
\begin{align*}
\text{min } & x_1^2 - x_2^2, \\
2x_1 - x_2 & - \varepsilon \leq 0, \\
-2x_1 - x_2 & - \varepsilon \leq 0.
\end{align*}
\]

At \( \varepsilon = 0 \), the problem still reduces to (20). The tangent quadratic problem

\[
\begin{align*}
\text{min } & d_1^2 - d_2^2, \\
2d_1 - d_2 & - 1 \leq 0, \\
-2d_1 - d_2 & - 1 \leq 0.
\end{align*}
\]

The point \( d = (0, 0)' \) is obviously a stationary point of this problem. We prove that it is not a pseudo-solution: \( f(d) = 0 \); hence, by (12) the only subpartition that could be associated with \( d \) is \( (X_1, X_2) = (\emptyset, \emptyset) \). But the Hessian of the Lagrangian is not positive; hence, \((\emptyset, \emptyset)\) cannot be a pseudo-optimal subpartition.

References


A Minimax Theorem for Vector-Valued Functions

P. FCMR̃

Communicated by P. L. Yu

Abstract. In this work, as usual in vector-valued optimization, we consider the partial ordering induced in a topological vector space by a closed and convex cone. In this way, we define maximal and minimal sets of a vector-valued function and consider minimax problems in this setting. Under suitable hypotheses (continuity, compactness, and special types of convexity), we prove that, for every

\[
\alpha \in \bigcup_{\gamma} \operatorname{Min}_{\gamma} f(x, y),
\]

there exists

\[
\beta \in \bigcup_{\gamma} \operatorname{Max}_{\gamma} f(x, y),
\]

such that \( \beta < \alpha \) (the exact meanings of the symbols are given in Section 2).

Key Words. Minimax theorems, vector-valued optimization.

1. Introduction

Minimax theorems for real-valued (or extended real-valued) functions

\[
f : X \times Y \rightarrow \mathbb{R} \quad \text{or} \quad \mathbb{R} \cup \{\pm \infty\}
\]

cite that, under suitable hypotheses of compactness, convexity, and continuity, the equality

\[
\inf_{\alpha \in X} \sup_{\beta \in Y} f(\alpha, \beta) = \sup_{\beta \in Y} \inf_{\alpha \in X} f(\alpha, \beta)
\]

holds. References 1-4 discuss this subject. See also Ref. 5 for a survey and extensive bibliographical references. The numerous studies of vector-valued optimization in recent years (e.g., see Refs. 6-9) seem to lead, in a natural way, to the investigation of minimax problems in this more general

*Professore Ordinario, Dipartimento di Matematica, Università di Genova, Genova, Italy.