ACTA MATHEMATICA VIETNAMICA
Volume 29, Number 2, 2004, pp. 141-148

ON A NECESSARY OPTIMALITY CONDITION WITH INVEXITY

HOANG XUAN PHU

Abstract. N. X. Ha and D. V. Luu presented in *Bull. Austral. Math. Soc.* **68** (2003) some sufficient conditions for invexity and a necessary optimality condition, in which invexity is used for ensuring the positivity of the Lagrange multiplier λ corresponding to the objective function. We show that the necessary optimality condition is wrong and invexity is misplaced there. Although $\lambda > 0$ follows immediately from the so-called Jourani constraint qualification and no additional explanation is needed, they spent almost three pages to prove the sufficiency of this condition for an invexity property, and then almost one page to derive $\lambda > 0$ from the invexity property and an additional condition of Slater type, while the latter ones are strong enough to yield the Jourani constraint qualification. Moreover, the other two main sufficient conditions for invexity in the mentioned paper yield the Jourani constraint qualification.

1. Introduction

Let X be a Banach space and $C \subset X$. Let f, g_i, $i \in I = \{1, 2, ..., l\}$, and h_j, $j \in J = \{1, 2, ..., m\}$, be real-valued functions on X, which are locally Lipschitz in C. Consider the problem

$$(P_1) \qquad \begin{cases} \text{minimize} \quad f(x), \\ \text{subject to} \\ g_i(x) \le 0, \quad i \in I, \\ h_j(x) = 0, \quad j \in J, \\ x \in C. \end{cases}$$

The following Lagrange multiplier rule is one of the most elegant necessary optimality conditions for this problem.

Theorem 1. (Clarke [1], pp. 228–229) *Let x_* be a local minimizer of (P_1). Then there exist $\lambda \ge 0$, $\mu_i \ge 0$, $i \in I$, and $\nu_j \in \mathbb{R}$, $j \in J$, not all zero, such that*

Received September 3, 2003.
2000 *Mathematics Subject Classification.* Primary 26B25, 49K27, Secondary 90C30.
Key words and phrases. Sufficient conditions for invexity, necessary optimality condition, Lagrange multiplier rule.

$$(1)\qquad \begin{aligned} &0 \in \lambda \partial f(x_*) + \sum_{i \in I} \mu_i \partial g_i(x_*) + \sum_{j \in J} \nu_j \partial h_j(x_*) + N_C(x_*), \\ &\mu_i g_i(x_*) = 0, \quad i \in I. \end{aligned}$$

Recall (see Clarke [1]) that for a function f which is Lipschitz near a given $x \in X$, its *generalized directional derivative* at x in direction $v \in X$ is defined by

$$f^\circ(x; v) = \limsup_{y \to x,\, t \downarrow 0} \frac{f(y + tv) - f(y)}{t}$$

and its *generalized gradient* at x is

$$\partial f(x) = \{\xi \in X^* : f^\circ(x; v) \geq \langle \xi, v \rangle \text{ for all } v \in X\}.$$

For a nonempty subset C of X, denote by

$$\begin{aligned} d_C(x) &= \inf\{\|x - c\| : c \in C\}, \\ T_C(x) &= \{v \in X : d_C^\circ(x; v) = 0\}, \\ N_C(x) &= \{\xi \in X^* : \langle \xi, v \rangle \leq 0 \text{ for all } v \in T_C(x)\} \end{aligned}$$

the *distance function*, the *tangent cone*, and the *normal cone* to C at $x \in X$, respectively.

There are two quite natural questions:

(i) When is the Lagrange multiplier λ corresponding to the objective function f definitely positive (i.e., $\lambda > 0$) so that one can put $\lambda = 1$?

(ii) When is such a necessary optimality condition as given in Theorem 1 sufficient for a global or a local minimum?

A classical pattern for answering these questions is the Kuhn-Tucker theorem dealing with problems without equality constraints. It ensures $\lambda > 0$ when the convexity of all constraint functions and a Slater condition are assumed. If the objective function and all constraint functions and the set C are convex, and if some feasible point x_* satisfies the necessary optimality condition in Theorem 1 for $\lambda = 1$, then x_* is a global minimizer (see [5], p. 68).

It is a traditional idea to replace the above mentioned convexity assumptions by some suitable kinds of generalized convexities, which had already been done repeatedly. Note that (generalized) convexities are not definitely necessary for these purposes. To show $\lambda > 0$, one actually needs a weaker regularity condition for the derivatives of constraint functions, and convexity and some kind of its generalizations are just used to imply this condition (see [5], p. 74).

Applying Theorem 1, Ha and Luu [3] proved a necessary optimality condition, in which invexity is used instead of convexity to ensure $\lambda > 0$. The main critical points are:

(α) This modified necessary optimality condition is backward and wrong.

(β) Invexity does not provide any proper advantage in this context and it is misplaced there.

We will explain (α) and (β) in Sections 2 and 3, respectively.

2. A wrong necessary optimality condition

Let us recall the corresponding result of Ha and Luu first.

Theorem 2. ([3], Theorem 5.1) *Let x_* be a local minimizer of (P_1). Then:*

(a) *There exist $\lambda \geq 0$, $\mu_i \geq 0$, $i \in I$, and $\nu_j \in \mathbb{R}$, $j \in J$, not all zero, such that for all $d \in T_C(x_*)$,*

$$\begin{aligned} &\lambda f^\circ(x_*; d) + \sum_{i \in I_0} \mu_i g_i^\circ(x_*; d) + \sum_{j \in J} \nu_j h_j^\circ(x_*; d) \geq 0, \\ &\mu_i g_i(x_*) = 0, \quad i \in I, \end{aligned} \tag{2}$$

where $I_0 = \{i \in I : g_i(x_*) = 0\}$.

(b) *$\lambda > 0$ holds definitely if the following are fulfilled:*

- *The functions g_i ($i \in I_0$) and h_j ($j \in J$) have the invexity property: there exists a map $\omega : X \to T_C(x_*)$ such that, for every $x \in X$,*

$$\begin{aligned} g_i(x) - g_i(x_*) &\geq g_i^\circ(x_*; \omega(x)), \quad \forall\, i \in I_0, \\ h_j(x) - h_j(x_*) &= h_j^\circ(x_*; \omega(x)), \quad \forall\, j \in J. \end{aligned} \tag{3}$$

- *For every $(\mu_{I_0}, \nu) \in \left(\mathbb{R}_+^{|I_0|} \times \mathbb{R}^m\right) \setminus \{0\}$ there exists $x \in X$ such that*

$$\langle \mu_{I_0}, G_{I_0}(x) \rangle + \langle \nu, H(x) \rangle < 0, \tag{4}$$

where $\mu_{I_0} = (\mu_i)_{i \in I_0}$, $\nu = (\nu_j)_{j \in J}$, $|I_0|$ is the cardinal number of I_0, $\mathbb{R}_+^{|I_0|}$ is the positive orthant of $\mathbb{R}^{|I_0|}$, $G_{I_0} = \big(g_i\big)_{i \in I_0}$, and $H = (h_1, ..., h_m)$.

Note that the invexity mentioned in (3) is originated to Reiland [13] who generalized the concept of Hanson [4] for Lipschitz functions.

Normally, one derives necessary optimality conditions using generalized gradients from some conditions using directional derivatives, not reversely, as done in Theorem 2. This work is not only strange, but also dangerous. Indeed, Theorem 2 is wrong, as shown in the following.

Counter-example 1. Let $X = C = \mathbb{R}$, $l = m = 1$, $f(x) = x$, $g_1(x) = -1$, and

$$h_1(x) = \begin{cases} 2x & \text{if } x < 0, \\ x & \text{if } x \geq 0. \end{cases}$$

Obviously, $x_* = 0$ is the unique feasible point of Problem (P_1), and therefore, it is the minimizer of (P_1). We have

$$\begin{aligned} &\partial f(0) = \{1\}, \quad f^\circ(0; -1) = -1, \quad f^\circ(0; 1) = 1, \\ &\partial h(0) = [1, 2], \quad h^\circ(0; -1) = -1, \quad h^\circ(0; 1) = 2. \end{aligned}$$

If there are $\lambda \geq 0$, $\mu_1 \geq 0$ and $\nu_1 \in \mathbb{R}$ such that (2) holds for all $d \in T_{\mathbb{R}}(0) = \mathbb{R}$, then $\mu_1 = 0$ and

$$-\lambda-\nu_1=\lambda f^\circ(0;-1)+\nu_1 h^\circ(0;-1)\geq 0,$$
$$\lambda+2\nu_1=\lambda f^\circ(0;1)+\nu_1 h^\circ(0;1)\geq 0.$$

The addition of two inequalities yields $\nu_1 \geq 0$. But it follows from $\lambda \geq 0$ and the first inequality that $\nu_1 \leq -\lambda \leq 0$. Thus $\nu_1 = 0$ and, therefore, $\lambda = 0$. Hence, all Lagrange multipliers are zero, a contradiction. Consequently, part (a) of Theorem 2 is false. Since part (b) depends on (a), it is also false. Thus, Theorem 2 (i.e., Theorem 5.1 of Ha and Luu in [3]) is wrong.

In the proof of the mentioned theorem, the authors used the relation

$$\max\{\nu_j\langle\xi,d\rangle:\xi\in\partial h_j(x_*)\}\leq\nu_j\max\{\langle\xi,d\rangle:\xi\in\partial h_j(x_*)\}=\nu_j h_j^\circ(x_*;d) \tag{5}$$

(for all $j \in J$ and $d \in T_C(x_*)$), which is true if $\nu_j \geq 0$. For $\nu_j < 0$,

$$\nu_j\max\{\langle\xi,d\rangle:\xi\in\partial h_j(x_*)\}=\min\{\nu_j\langle\xi,d\rangle:\xi\in\partial h_j(x_*)\};$$

so (5) is true if and only if

$$\{\langle\xi,d\rangle:\xi\in\partial h_j(x_*)\}\text{ is a singleton.}$$

Hence, the mentioned error can be avoided and part (a) of Theorem 2 can be saved only in some special cases, namely:

$$J=\emptyset, \tag{6}$$

or, for all $j \in J$,

$$\partial h_j(x_*)\text{ is a singleton,} \tag{7}$$

or, even weaker,

$$\text{for all } d\in T_C(x_*),\ \ \{\langle\xi,d\rangle:\xi\in\partial h_j(x_*)\}\text{ is a singleton,} \tag{8}$$

or

$$\partial h_j(x_*)\text{ is symmetric, i.e., }\ \partial h_j(x_*)=-\partial h_j(x_*). \tag{9}$$

In fact, if (9) is fulfilled, then ν_j in (1), and therefore in (2), can be assumed to be non-negative because $\nu_j\partial h_j(x_*) = -\nu_j\partial h_j(x_*)$. In such a way, (5) becomes true. But after such a specification, the necessary condition in Theorem 5.1 [3] is much more weaker than the original one in Theorem 1. Then, why had it been derived? Just to demonstrate the use of invexity?

3. Misplaced invexity

Assuming that the first part of Theorem 2 had been corrected, we now discuss the second part saying when $\lambda > 0$. An important condition for it is that the functions g_i ($i \in I_0$) and h_j ($j \in J$) have the property (3). 9 pages in [3] are devoted to proving some sufficient conditions for this invexity property. One of them is the generalized Mangasarian-Fromovitz constraint qualification (see [9] and [10]) stated for continuously Fréchet differentiable functions $h_1, h_2, ..., h_m$ as follows:

$$(10)\qquad \begin{cases} \text{there exists } d_0 \in \operatorname{int} T_C(x_*) \text{ such that} \\ \langle h_j'(x_*), d_0\rangle = 0 \quad \text{for all} \quad j \in J, \\ \langle \xi_i, d_0\rangle < 0 \quad \text{for all } \xi_i \in \partial g_i(x_*) \text{ and } i \in I_0, \quad \text{and} \\ h_1'(x_*), h_2'(x_*), ..., h_m'(x_*) \;\; \text{are linearly independent.} \end{cases}$$

Another sufficient condition for (3) is the regularity condition of Robinson type [14]

$$(11)\qquad 0 \in \operatorname{int} \Big(F^\circ\big(T_C(x_*)\big) + \mathbb{R}_+^{|I_0|} \times \{0^m\} \Big),$$

where $F^\circ = (G^\circ_{I_0}, H^\circ)$, $G^\circ_{I_0} = \big(g_i^\circ(x_*;.)\big)_{i \in I_0}$, $H^\circ = \big(h_1^\circ(x_*;.), ..., h_m^\circ(x_*;.)\big)$, and 0^m is the origin of $\mathbb{R}^m$. The third condition is the so-called Jourani constraint qualification [6]

(12)
$$0 \notin \sum_{i \in I_0} \mu_i \partial g_i(x_*) + \sum_{j \in J} \nu_j \partial h_j(x_*) + N_C(x_*) \;\text{ for all } (\mu_{I_0}, \nu) \in \Big(\mathbb{R}_+^{|I_0|} \times \mathbb{R}^m\Big) \setminus \{0\}.$$

To evaluate the results of [3], let us clarify the relations between these three conditions.

Proposition 1. *If* $h_1, h_2, ..., h_m$ *are continuously Fréchet differentiable then* (10) *implies* (12).

Proof. Assume that (10) is true but (12) fails, then

$$\sum_{i \in I_0} \mu_i \xi_i + \sum_{j \in J} \nu_j h_j'(x_*) + \eta = 0$$

for some $(\mu_{I_0}, \nu) \in \Big(\mathbb{R}_+^{|I_0|} \times \mathbb{R}^m\Big) \setminus \{0\}$, $\xi_i \in \partial g_i(x_*)$, $i \in I_0$, and $\eta \in N_C(x_*)$. Since $\langle h_j'(x_*), d_0\rangle = 0$ for all $j \in J$,

$$\sum_{i \in I_0} \mu_i \langle \xi_i, d_0\rangle + \langle \eta, d_0\rangle = 0.$$

As $d_0 \in \operatorname{int} T_C(x_*)$ and $\langle \eta, d\rangle \le 0$ for all $d \in T_C(x_*)$, we have $\langle \eta, d_0\rangle < 0$ whenever $\eta \neq 0$. Therefore, it follows from $\langle \xi_i, d_0\rangle < 0$ for all $\xi_i \in \partial g_i(x_*))$ that $\mu_i = 0$ for all $i \in I_0$ and $\eta = 0$. Hence, $\sum_{j \in J} \nu_j h_j'(x_*) = 0$, which yields $\nu_1 = ... = \nu_m = 0$ because $h_1'(x_*), ..., h_m'(x_*)$ are linearly independent. Thus, all μ_i and ν_j are zero. This contradiction shows that (10) implies (12). □

Note that Jourani [6] already showed the equivalence of (10) and (12) when X is finite dimensional.

Proposition 2. *Assume* $J = \emptyset$ *or, for all* $j \in J$, (7) *or* (8) *or* (9) *is satisfied. Then* (11) *implies* (12).

Proof. Assume that (12) is false, i.e.,

$$0 \in \sum_{i\in I_0} \mu_i \partial g_i(x_*) + \sum_{j\in J} \nu_j \partial h_j(x_*) + N_C(x_*)$$

for some $(\mu_{I_0}, \nu) \in \left(\mathbb{R}_+^{|I_0|} \times \mathbb{R}^m\right) \setminus \{0\}$, where $\nu_j \geq 0$ if (9) is fulfilled. Then

$$\sum_{i\in I_0} \mu_i \vartheta_i + \sum_{j\in J} \nu_j \xi_j + \eta = 0$$

for some $\vartheta_i \in \partial g_i(x_*)$, $i \in I_0$, $\xi_j \in \partial h_j(x_*)$, $j \in J$, and $\eta \in N_C(x_*)$. By applying

$$f^\circ(x; d) = \max\{\langle \xi, d\rangle : \xi \in \partial f(x)\}$$

and

$$\langle \xi, d\rangle \leq 0 \quad \text{for all } \ \xi \in N_C(x),\ d \in T_C(x),$$

and by assumptions, we have

$$\sum_{i\in I_0} \mu_i g_i^\circ(x_*; d) + \sum_{j\in J} \nu_j h_j^\circ(x_*; d) \geq 0 \ \text{ for all } d \in T_C(x_*). \tag{13}$$

On the other hand, if (11) holds then

$$F^\circ\big(T_C(x_*)\big) + \mathbb{R}_+^{|I_0|} \times \{0^m\} = \mathbb{R}^{|I_0|+m}, \tag{14}$$

because $F^\circ\big(T_C(x_*)\big) + \mathbb{R}^{|I_0|} \times \{0^m\}$ is a cone. Since $(\mu_{I_0}, \nu) \in \left(\mathbb{R}_+^{|I_0|} \times \mathbb{R}^m\right) \setminus \{0\}$, at least one of μ_i, $i \in I_0$, and ν_j, $j \in J$, must be different from zero. Consequently, by (14), there exists a $d \in T_C(x_*)$ satisfying

$$\sum_{i\in I_0} \mu_i g_i^\circ(x_*; d) + \sum_{j\in J} \nu_j h_j^\circ(x_*; d) < 0,$$

which conflicts with (13). We have shown that, under the assumptions of our proposition, (11) implies (12). □

In general, if (7) or (8) or (9) is not assumed, then (11) does not imply (12), as shown in the following.

Example 2. Let $X = \mathbb{R}^2$, $C = \mathbb{R}_+^2$, $I = \emptyset$, $J = \{1\}$, and $h_1(x_1, x_2) = -x_1 + |x_2|$. For $x_* = (0, 0)$, we have $T_C(x_*) = \mathbb{R}_+^2$. Therefore, it follows from

$$h_1^\circ(x_*; d) = \begin{cases} -1 & \text{if } d = (1, 0) \\ +1 & \text{if } d = (0, 1), \end{cases}$$

that $F^\circ(T_C(x_*)) = h_1^\circ(x_*; T_C(x_*)) = \mathbb{R}$, i.e., (11) is true. But (12) does not hold for $\nu_1 = -1$, because the formulas $\partial h_1(x_*) = \{(-1, \rho) : |\rho| \leq 1\}$ and $N_C(x_*) = -\mathbb{R}_+^2$ imply

$$0 \in -\partial h_1(x_*) + N_C(x_*).$$

This example shows that, in general, (11) does not imply that the Lagrange multiplier λ in Theorem 1 is positive, although condition (4) is satisfied.

Due to Propositions 1 and 2, it is too expensive to show separately and independently that each one of (10), (11), and (12) implies the invexity property (3), as it was done in [3].

In [3], all sufficient conditions for property (3) are needed for part (b) of Theorem 2, i.e., for showing $\lambda > 0$. But this is an unacceptable roundabout way. To see it, let us use an simple example for illustration.

Consider the necessary condition

$$\lambda f'(x_*) + \mu g'(x_*) = 0, \quad \lambda \geq 0, \quad \mu \geq 0, \quad \mu g(x_*) = 0, \quad (\lambda, \mu) \neq (0,0),$$

where f and g are real-valued functions on $\mathbb{R}$ which are continuously differentiable near x_*. If $g'(x_*) \neq 0$ is assumed, then it yields immediately that $\lambda > 0$, and there is nothing to prove anymore. But, following [3], one would do similarly as follows. $g'(x_*) \neq 0$ implies $g'(x) \neq 0$ in a neighborhood of x_*, i.e., g has no stationary point in this neighborhood, and therefore, it is invex there. Since this invexity is not enough, one must assume, in addition, that there exists an x satisfying $\mu g(x) < 0$, and comes to the conclusion $\mu g'(x_*) \neq 0$, which yields $\lambda > 0$.

It is unbelievable, but it was really done in a similar way. In fact, the disadvantage is much more. Obviously, (12) yields immediately that the Lagrange multiplier λ in Theorem 1 is positive. Since Theorem 2 was derived from Theorem 1, the Lagrange multiplier λ in Theorem 2 is also positive (if Theorem 2 is true). There is nothing to explain else. But almost three pages are needed in [3] to prove the sufficiency of (12) for the invexity property (3), and then almost one page is used to derive $\lambda > 0$ from (3) and (4). The same work amount is spent to come from condition (10) to $\lambda > 0$, while Proposition 1, whose proof needs at most ten lines, is completely sufficient for this purpose.

Due to the above fact, the existence of Theorem 2 is eligible only if the conditions (3) and (4) together are relatively weak. But, unfortunately, they are strong enough to yield (12), as shown in the following.

Proposition 3. *Assume $J = \emptyset$ or, for all $j \in J$,* (7) *or* (8) *or* (9) *is satisfied. Then* (3) *and* (4) *imply* (12).

Proof. If (12) fails to be true, then, due to first part of the proof of Proposition 2, there exists $(\mu_{I_0}, \nu) \in \left(\mathbb{R}_+^{|I_0|} \times \mathbb{R}^m\right) \setminus \{0\}$ such that (13) holds, i.e.,

$$\sum_{i \in I_0} \mu_i g_i^\circ(x_*; d) + \sum_{j \in J} \nu_i h_j^\circ(x_*; d) \geq 0 \quad \text{for all } d \in T_C(x_*).$$

Take an x satisfying (4) and apply (3) for this x, it follows by addition of the corresponding inequalities and equalities that

$$\sum_{i \in I_0} \mu_i g_i^\circ(x_*; \omega(x)) + \sum_{j \in J} \nu_i h_j^\circ(x_*; \omega(x)) \leq \sum_{i \in I_0} \mu_i g_i(x) + \sum_{j \in J} \nu_i h_j(x) < 0,$$

a contradiction because $\omega(x) \in T_C(x_*)$. Hence, (3) and (4) yield (12). □

Note that the assumptions in Propositions 2 and 3 do not restrict our consideration, because (6) or (7) or (8) or (9) must be assumed, in order to correct the first part of Theorem 2 (see the remark at the end of Section 2).

Altogether, we see that invexity is misplaced in [3].

4. Concluding remarks

Although the title is “Sufficient conditions for invexity”, Theorem 5.1 [3] must be considered as the key result of that paper, because it is the only given use of the stated sufficient conditions for invexity, and another application for them is not visible (see corresponding comments in [11]). But Theorem 5.1 [3] is wrong, as shown in Section 2, and the sufficient conditions for invexity in [3] are no eligible tools for ensuring $\lambda > 0$ in Theorem 5.1 [3], as shown in Section 3.

Further essential errors combined with invexity and its generalizations in [2], [7], and [8] are analyzed in [11] and [12].

Acknowledgment

The author thanks Prof. Dr. Nguyen Dong Yen, Prof. Dr. Pham Huu Sach, and the referee for their valuable comments and suggestions.

References

[1] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley & Sons, New York, 1983.
[2] N. X. Ha and D. V. Luu, *Invexity of supremum and infimum functions*, Bull. Austral. Math. Soc. **65** (2002), 289–306.
[3] N. X. Ha and D. V. Luu, *Sufficient conditions for invexity*, Bull. Austral. Math. Soc. **68** (2003), 113–125.
[4] M. A. Hanson, *On sufficiency of the Kuhn-Tucker conditions*, J. Math. Anal. Appl. **80** (1981), 545–550.
[5] A. D. Ioffe and V. M. Tichomirov, *Theory of Extremal Problems*, North-Holland Publishing Company, Amsterdam, 1979.
[6] A. Jourani, *Constraint qualifications and Lagrange multipliers in nondifferentiable programming problems*, J. Optim. Theory Appl. **81** (1994), 533–548.
[7] D. V. Luu and N. X. Ha, *An invariant property of invex functions and applications*, Acta Math. Vietnam. **25** (2000), 181–193.
[8] D. V. Luu and P. T. Kien, *Sufficient optimality conditions under invexity hypotheses*, Vietnam J. Math. **28** (2000), 227–236.
[9] O. L. Magasarian and S. Fromovitz, *The Fritz-John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl. **17** (1967), 37–47.
[10] V. H. Nguyen, J. J. Strodiot, and R. Mifflin, *On conditions to have bounded multipliers in locally Lipschitz programming*, Math. Program. **18** (1980), 100–106.
[11] H. X. Phu, *Is invexity weaker than convexity?* Vietnam J. Math. **32** (2004), 87–94.
[12] H. X. Phu, *On some badly-solved problems with invexity.* Acta Math. Vietnam. **29** (2004), 89-106.
[13] T. W. Reiland, *Nonsmooth invexity*, Bull. Austral. Math. Soc. **42** (1990), 437–446.
[14] S. M. Robinson, *Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems*, SIAM J. Numer. Anal. **13** (1976), 479–513.

Institute of Mathematics
18 Hoang Quoc Viet Road, 10307 Hanoi, Vietnam.