Journal of Global Optimization (2006) 34: 535–549
DOI 10.1007/s10898-005-1650-5

Necessary Optimality Conditions for Bilevel Optimization Problems Using Convexificators

H. BABAHADDA and N. GADHI
Department of Mathematics, Dhar El Mehrez, Sidi Mohamed Ben Abdellah University, Fes, Morocco (e-mail: Hicham.Babahadda@gerad.ca, ngadhi@math.net)

(Received 14 August 2003; accepted in revised form 27 July 2005)

Abstract. In this work, we use a notion of convexificator (Jeyakumar, V. and Luc, D.T. (1999), Journal of Optimization Theory and Applicatons, 101, 599–621.) to establish necessary optimality conditions for bilevel optimization problems. For this end, we introduce an appropriate regularity condition to help us discern the Lagrange–Kuhn–Tucker multipliers.

Mathematics Subject Classifications (1991). Primary 49J52, Secondary 49K99.

Key words: Bilevel optimization, convexificator, continuous functions, Lagrange–Kuhn–Tucker multipliers, necessary optimality conditions, regularity condition

1. Introduction

In recent years, a great deal of research in nonsmooth analysis has focused on the development of generalized subdifferentials that provide sharp extremality conditions and good calculus rules for nonsmooth functions [4, 7, 10, 24, 27, 29]. Very recently, the idea of convexificators has been used to extend, unify, and sharpen various results in nonsmooth analysis and optimization [10, 15, 16]. In [17], Jeyakumar and Luc gave a revised version of convexificators by introducing the notion of a convexificator which is a closed set but is not necessarily bounded or convex. Such a new notion allows applications of convexificators to continuous functions.

The problem (P) considered in this paper is a sequence of two optimization problems in which the feasible region of the upper-level problem (P_1) is determined implicitly by the solution set of the lower-level problem (P_2). It may be given as follows

$$(P_1): \quad \begin{cases} \text{Minimize } F(x,y) \\ \text{subject to:} G(x,y) \leqslant 0,\ y \in S(x), \end{cases}$$

where, for each $x \in X$, $S(x)$ is the solution set of the following parametric optimization problem

$$(P_2): \quad \begin{cases} \text{Minimize} f(x,y) \\ \text{subject to:} g(x,y) \leqslant 0, \end{cases}$$

where $F, f\colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \longrightarrow \mathbb{R}, G = \left(G_1, \ldots, G_{m_2}\right)\colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \longrightarrow \mathbb{R}^{m_2}$ and $g = (g_1, \ldots, g_{m_1})\colon \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \longrightarrow \mathbb{R}^{m_1}$ are given; n_i and $m_i, i = 1, 2,$ are integers with $n_i \geqslant 1$ and $m_i \geqslant 0$.

We agree that, whenever $m_1 = 0$ or $m_2 = 0$, this means that the corresponding inequality constraint is absent in the (P). A pair $(\overline{x}, \overline{y})$ is said to be optimal solution to the (P) if it is an optimal solution to the following problem: $\min_{(x,y)\in\overline{S}} F(x, y)$ where

$$\overline{S} = \left\{(x, y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : G(x, y) \leqslant 0 \quad \text{and} \quad y \in S(x)\right\}.$$

A lot of research has been carried out in bilevel optimization problems [3, 6, 8, 13, 23, 25, 26, 30–32]. Ye and Zhu [31] give optimality conditions without convexity assumption on the lower level problem and without the assumption that the solution set $S(x)$ is a singleton. Under semi-Lipschitz property, Zhang [32] extends the classical approach to allow the nonsmooth problem data; he derives existence and optimality conditions for problems in terms of a graph set of the solution multifunction to the lower-level problem.

In this note, our approach consists in using an appropriate regularity condition and the notion of convexificator to detect necessary optimality conditions in terms of Lagrange–Kuhn–Tucker multipliers. Some examples that illustrate the usefulness of convexificators are also given. For a locally Lipschitz function, most known subdifferentials such as the subdifferential of Clarke, Michel–Penot, Ioffe–Mordukhovich and Treiman are convexificators. For more details, see [17] and the references therein.

The rest of the paper is written as follows: Section 2 contains basic definitions and preliminary results. Section 3 is devoted to the optimality conditions.

2. Preliminaries

Let $f\colon \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ be an extended real valued function. The expressions

$$f_d^-(x, v) := \liminf_{t \searrow 0} \, [f(x + tv) - f(x)]/t,$$
$$f_d^+(x, v) := \limsup_{t \searrow 0} \, [f(x + tv) - f(x)]/t$$

signify, respectively, the lower and upper Dini directional derivatives of f at x in the direction of v.

DEFINITION 1. [17]. The function $f\colon \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ is said to have a convexificator $\partial^* f(x)$ at x if $\partial^* f(x) \subset \mathbb{R}^p$ is closed and, for each $v \in \mathbb{R}^p$,

$$f_d^-(x, v) \leqslant \sup_{x^* \in \partial^* f(x)} \left\langle x^*, v\right\rangle \quad \text{and} \quad f_d^+(x, v) \geqslant \inf_{x^* \in \partial^* f(x)} \left\langle x^*, v\right\rangle.$$

Note that convexificators are not necessarily compact or convex [10]. These relaxations allow applications to a large class of nonsmooth continuous functions. For instance, the function $f:\mathbb{R}\to\mathbb{R}$ defined by

$$f(x)=\begin{cases}\sqrt{x} & \text{if } x\geqslant 0,\\ -\sqrt{-x} & \text{if } x<0,\end{cases}$$

admits noncompact convexificators at 0 of the form $[\alpha,\infty)$ with $\alpha\in\mathbb{R}$. On the other hand, the function $f:\mathbb{R}\to\mathbb{R}$, defined by

$$f(x)=-|x|,$$

admits a nonconvex convexificator $\partial^* f(0)=\{1,-1\}$ at 0.

Denoting by $\partial^\diamond f(.)$ and $\partial^\circ f(.)$ the Michel–Penot subdifferential and the Clarke generalized subdifferential, we have the following remarks and examples.

Remark 1. Let $f:\mathbb{R}^p\to\mathbb{R}\cup\{+\infty\}$ be finite at a point $\overline{x}\in\mathbb{R}^p$.
If f is locally Lipschitz at $\overline{x}$, then $\partial^\circ f(\overline{x})$ and $\partial^\diamond f(\overline{x})$ are convexificators of f at $\overline{x}$. However, the convex hull of a convexificator of a locally Lipschitz function may be strictly contained in both the Clarke and Michel–Penot subdifferentials.

Here, the set $\partial^\circ f(x)$ designes the Clarke generalized gradient of f at x; i.e.

$$\partial^\circ f(x):=\left\{x^*\in\mathbb{R}^p:\ \limsup_{u\to x,\ t\searrow 0}\frac{f(u+tv)-f(u)}{t}\geqslant\langle x^*,v\rangle\quad\forall v\in\mathbb{R}^p\right\}$$

and $\partial^\diamond f(x)$ designes the Michel–Penot subdifferential of f at x; i.e.

$$\partial^\diamond f(x)=\left\{x^*\in\mathbb{R}^p: x^*\leqslant f^\diamond(x,.)\right\}$$

where

$$f^\diamond(x,v)=\sup_{w\in\mathbb{R}^p}\limsup_{t\to 0^+} t^{-1}[f(x+tw+tv)-f(x+tw)].$$

for more details, see [21, 22].

EXAMPLE 1. [17]. Define $f:\mathbb{R}^2\to\mathbb{R}$ by

$$f(x)=|x|-|y|.$$

It can easily be verified that

$$\partial^* f\ (0) = \{(1,-1),(-1,1)\}$$

is a convexificator of f at 0, whereas

$$\partial^\diamond f\,(0) = \partial^\circ f\,(0) = \mathrm{co}(\{(1,1), \quad (-1,1), \quad (1,-1), \quad (-1,-1)\}).$$

It is clear that

$$\mathrm{co}\left(\partial^* f\ (0)\right) \subset \partial^\diamond f\ (0) = \partial^\circ f\ (0)\,.$$

Clearly, this example shows that certain results such as the necessary optimality conditions that are expressed in terms of $\partial^* f\ (x)$ may provide sharp conditions even for locally Lipschitz functions.

To progress, we need the following definition.

DEFINITION 2. [17]. A set valued mapping $F\colon \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ is upper semicontinuous (u.s.c.) at x, if for each $\varepsilon > 0$, there exists $\delta > 0$ such that, for each $x' \in x + \delta \mathbb{B}_{\mathbb{R}^p}$,

$$F\left(x'\right) \subset F\,(x) + \varepsilon \mathbb{B}_{\mathbb{R}^q},$$

where $\mathbb{B}_{\mathbb{R}^p}$ and $\mathbb{B}_{\mathbb{R}^q}$ are the unit balls in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively.

In order to give an example of non locally Lipschitz function, let us recall the following definition.

DEFINITION 3. [24]. Let $f\colon \mathbb{R}^p \to \overline{\mathbb{R}} := [-\infty, +\infty]$ be an extended-real-valued function and $\overline{x} \in \mathrm{dom}(f)$. The symmetric subdifferential of f at $\overline{x}$ is defined by

$$\partial^0 f\ (\overline{x}) := \partial f\ (\overline{x}) \sqcup [-\partial(-f)\,(\overline{x})]$$

where $\partial f\ (\overline{x}) := \lim_{x \xrightarrow{f} \overline{x},} \sup_{\varepsilon \searrow 0} \widehat{\partial}_\varepsilon f\ (x)$ and $\widehat{\partial}_\varepsilon f\ (x)$ is the $\varepsilon-$ Fréchet subdifferential of f at x. For more details see [24].

Note that sufficient conditions for the upper semicontinuity of $\partial^0 f\ (.)$ can be found in [14] and [19].

PROPOSITION 1. *Let* $f\colon \mathbb{R}^p \to \overline{\mathbb{R}} := [-\infty, +\infty]$ *be continuous and* $\overline{x} \in$ *dom* (f). *Suppose that* $\partial^0 f\ (\overline{x})$ *is closed and that* $\partial^0 f\ (.)$ *is upper semicontinuous at* $\overline{x}$. *Then* $\partial^0 f\ (\overline{x})$ *is a convexificator of* f *at* $\overline{x}$.

Proof. Let $\varepsilon > 0$. By the upper semicontinuity of $\partial^0 f\,(.)$, there exists $\delta > 0$ such that

$$\partial^0 f\,\left(x'\right) \subset \partial^0 f\,(\overline{x}) + \varepsilon \mathbb{B}_{\mathbb{R}^q},$$

for all $x \in \overline{x} + \delta \mathbb{B}_{\mathbb{R}^p}$.
Using Theorem 2.3 of [16] (the mean value theorem), there exists $c \in]x, \overline{x}[$ such that

$$f\,(x) - f\,(\overline{x}) \in \partial^0 f\,(c)\,(x - \overline{x}) \subset \partial^0 f\,(\overline{x})\,(x - \overline{x}) + \varepsilon\,\|x - \overline{x}\|\,\mathbb{B}_{\mathbb{R}}.$$

Now, let $v \in \mathbb{R}^p$. Since $\mathbb{B}_{\mathbb{R}}$ is compact,

$$f_d^-\,(\overline{x}, v) \in \partial^0 f\,(\overline{x})\,(v) + \varepsilon\,\|v\|\,\mathbb{B}_{\mathbb{R}} \quad \text{and} \quad f_d^+\,(\overline{x}, v) \in \partial^0 f\,(\overline{x})\,(v) + \varepsilon\,\|v\|\,\mathbb{B}_{\mathbb{R}}.$$

Consequently, there exist $x_1^*,\, x_2^* \in \partial^0 f\,(\overline{x})$ and $b_1,\, b_2 \in \mathbb{B}_{\mathbb{R}}$ such that

$$f_d^-\,(\overline{x}, v) = \left\langle x_1^*, v\right\rangle + \varepsilon\,\|v\|\,b_1 \quad \text{and} \quad f_d^+\,(\overline{x}, v) = \left\langle x_2^*, v\right\rangle + \varepsilon\,\|v\|\,b_2.$$

Then,

$$f_d^-\,(\overline{x}, v) \leqslant \sup_{x^* \in \partial^0 f\,(\overline{x})} \left\langle x^*, v\right\rangle + \varepsilon\,\|v\| \quad \text{and} \quad f_d^+\,(\overline{x}, v) \geqslant \inf_{x^* \in \partial^0 f\,(\overline{x})} \left\langle x^*, v\right\rangle - \varepsilon\,\|v\|\,.$$

Letting $\varepsilon \to 0$, one gets

$$f_d^-\,(\overline{x}, v) \leqslant \sup_{x^* \in \partial^0 f\,(x)} \left\langle x^*, v\right\rangle \quad \text{and} \quad f_d^+\,(\overline{x}, v) \geqslant \inf_{x^* \in \partial^0 f\,(x)} \left\langle x^*, v\right\rangle.$$

The proof is finished. □

Remark 2. 1. It has been proved by Amahroq and Gadhi [2] that if f is continuous then $\partial^0 f\,(\overline{x})$ is an approximation of f at $\overline{x}$. (For more details, see [1, 2, 18])
2. Instead of using the symmetric subdifferential of f at $\overline{x}$, one can formulate the above result by means of the notion of approximation [1, 18].

Now, we recall the chain rule for composite functions in terms of convexificators established by Jeyakumar and Luc in [17].

PROPOSITION 2. [17]. *Let $f = (f_1, \ldots, f_n)$ be a continuous function from $\mathbb{R}^p$ to $\mathbb{R}^n$, and g be continuous function from $\mathbb{R}^n$ to $\mathbb{R}$. Suppose that, for each $i = 1, 2, \ldots, n$, f_i admits a bounded convexificator $\partial^* f_i\,(\overline{x})$ and that g admits* *a bounded convexificator* $\partial^* g(f(\overline{x}))$ *at* $f(\overline{x})$. *For each* $i = 1, \ldots, n$, *if* $\partial^* f_i$ *is u.s.c. at* $\overline{x}$ *and* $\partial^* g$ *is u.s.c. at* $f(\overline{x})$, *then the set*

$$\partial^*(g \circ f)(\overline{x}) := \partial^* g(f(\overline{x}))\left(\partial^* f_1(\overline{x}), \ldots, \partial^* f_n(\overline{x})\right)$$

is a convexificator of $g \circ f$ *at* $\overline{x}$.

Since the convex hull of a convexificator of a locally Lipschitz function may be strictly contained in the Clarke subdifferential, Corollary 3 is an extension of Proposition 2.3.12 [4].

COROLLARY 3. *Let* $f = (f_1, \ldots, f_n)$ *be a continuous function from* X *to* $\mathbb{R}^n$. *Suppose that for* $i = 1, 2, \ldots, n$, *the function* f_i *admits a bounded convexificator* $\partial^* f_i(\overline{x})$ *at* $\overline{x}$.
Let

$$h(x) = \max\{f_i(x) : i = 1, 2, \ldots, n\}$$

and $I(\overline{x}) = \{i : f_i(\overline{x}) = h(\overline{x})\}$. *Then* $\mathrm{co}\left\{\cup_{i \in I(\overline{x})} \partial^* f_i(\overline{x})\right\}$ *is a convexificator of* h *at* $\overline{x}$, *where* "co" *denotes the convex hull.*

Proof. The proof is a consequence of Proposition 2 by considering

$$f(x) := (f_1(x), f_2(x), \ldots, f_n(x))$$

and

$$g(t_1, t_2, \ldots, t_n) := \max(t_1, t_2, \ldots, t_n).$$

Since g is Lipschitz and convex,

$$\partial_c g(s_1, \ldots, s_n) = \left\{ (r_1, \ldots, r_n) : \begin{array}{c} r_i \geqslant 0, \quad \sum_{i=1}^{n} r_i = 1 \\ \text{and } r_i = 0 \text{ whenever } s_i < g(s_1, \ldots, s_n) \end{array} \right\},$$

is a convexificator for g at $(s_1, \ldots, s_n)$. Consequently $\mathrm{co}\{\partial^* f_i(\overline{x}) : i \in I(\overline{x})\}$ is a convexificator for h at $\overline{x}$.
The proof is thus complete. □

Similarly, we deduce the following result which is a variant of Theorem 2.8.2 [4]. See also [5, 28].

COROLLARY 4. *Let* $\mathbb{T}$ *be a sequentially compact space,* $\overline{x} \in \mathbb{R}^p$, $f_t : \mathbb{R}^p \to \overline{\mathbb{R}}$ *and*

$$h(x) = \sup_{t \in \mathbb{T}}\{f_t(x)\} \quad \text{and} \quad J(\overline{x}) = \{t \in \mathbb{T} : f_t(\overline{x}) = h(\overline{x})\}.$$

Suppose that there exists a neighborhood U of $\overline{x}$ in Y such that for each $t \in \mathbb{T}$, the function f_t is finite on U and admits a bounded convexificator on U. If in addition $t \longmapsto f_t$ is upper semicontinuous then, $\text{cl co}\{\partial^* f_t(\overline{x}) : t \in J(\overline{x})\}$ *is a convexificator of h at $\overline{x}$.*

Proof. It suffices to repeat (with very slight modification) the argument of the first part of the proof of Theorem 2.8.2 in Clarke [4]. □

Let $H\colon \mathbb{R}^p \rightrightarrows \mathbb{R}^q$ be a set-valued mapping. For every $y^* \in \mathbb{R}^q$, the support function of H at x is defined by

$$C_H\left(y^*, x\right) := \sup_{y \in H(x)} \langle y^*, y \rangle,$$

where $\langle ., . \rangle$ is the dual pairs.
Suppose that for all $x \in X$, $H(x)$ is a nonempty, closed and convex set. The distance function of H to zero,

$$d(0, H(x)) = \inf\{\|y\| : y \in H(x)\}$$

is related to the support function of H by the relation

$$d\,(0, H(x)) = \max_{y^* \in Y_H^* \cap \mathbb{R}^q} - C_H\left(y^*, x\right),$$

where Y_H^* denotes the barrier cone of H defined by

$$Y_H^* := \left\{ y^* \in \mathbb{R}^q\colon \sup_{y \in H(x)} \langle y^*, y \rangle < +\infty \right\}.$$

If $d\,(0, H(x)) > 0$ then there is a unique $y^* \in Y_H^* \cap \mathbb{B}_{\mathbb{R}^q}$ satisfying $\|y^*\| = 1$ and $d\,(0, H(x)) = -C_H\,(y^*, x)$, see [11] and [28].

Using Corollary 4, one can deduce the following result which is an extension of Proposition 2.2 [11].

COROLLARY 5. *Suppose that there exists a neighborhood $U \subset X$ of $\overline{x}$ such that for each $x \in U$ and $y^* \in Y_H^* \cap \mathbb{B}_{\mathbb{R}^q}$, the support function $C_H\,(y^*, .)$ is continuous on U and admit a bounded convexificator $\partial^* C_H\,(y^*, .)\,(x)$. Then, the distance function $d\,(0, H(x))$ admits*

$$\text{cl co}\left\{-\partial^* C_H\left(y^*, .\right)(x) : y^* \in J\,(x)\right\}$$

as a convexificator at x, where $J\,(x) = \left\{y^* \in Y_H^* : \|y^*\| \leqslant 1, d\,(y, H(x)) = -C_H(y^*, x)\right\}$.
If in addition, $d\,(0, H(x)) > 0$ then $J(x)$ consists of only one single element y^* with $\|y^*\| = 1$.

3. Optimality Conditions

For all the sequel, it is assumed that the leader presuppose cooperation of the follower in the sense that the latter will choose in each time that solution in $S(x)$ which is best suited with respect to the leader's objective function.

In this case, according to Stephane Dempe [9], (P) can be replaced by

$$(P^*): \begin{cases} \text{Minimize} & F(x,y) \\ & G(x,y) \leqslant 0,\ g(x,y) \leqslant 0, \\ \text{subject to:} & f(x,y) - V(x) \leqslant 0, \\ & (x,y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}, \end{cases}$$

provided that (P^*) has an optimal solution [20], where

$$V(x) := \min_y \left\{ f(x,y) : g(x,y) \leqslant 0,\ y \in \mathbb{R}^{n_2} \right\}.$$

Remark 3. Dempe [9] has proved that under the following hypotheses (H_1), (H_2) and (H_3), the optimization problem (P) has at least one optimal solution.

(H_1): $F(.,.)$, $f(.,.)$, $g(.,.)$ and $G(.,.)$ are lower semicontinuous (l.s.c.) on $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$;

(H_2): $V(.)$ is upper semicontinuous (u.s.c.) on $\mathbb{R}^{n_1}$;

(H_3): The problem (P^*) has at least one feasible solution (i.e., the infimal value v^* of the function $F(.,.)$ on the feasible set of this problem is less than infinity), there exists $v^* < c < \infty$ such that

$$M := \left\{ (x,y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} : G(x,y) \leqslant 0,\ g(x,y) \leqslant 0,\ F(x,y) \leqslant c \right\}$$

is not empty and bounded.

For more details, we refer the interested reader to [9].

The following regularity is in the line of Amahroq and Gadhi's constraint qualification [2]. We will use it to establish necessary optimality conditions in terms of Lagrange–Kuhn–Tucker multipliers.

DEFINITION 4. The problem (P) is said to be regular at $(\overline{x}, \overline{y})$ if there exists a neighborhood U of $(\overline{x}, \overline{y})$ and $\delta, \beta > 0$ such that :

$\forall (\mu, \upsilon) \in \mathbb{R}_+^{m_1} \times \mathbb{R}_+^{m_2}$, $\forall (x,y) \in U$, $\forall x_g^* \in \operatorname{co} \partial^* g(x,y)$, $\forall x_G^* \in \operatorname{co} \partial^* G(x,y)$, $\forall x_f^* \in \operatorname{co} \partial^* f(x,y)$, $\forall x_V^* \in \partial^* V(x) \times \{0\}$ $\exists \xi \in \delta \mathbb{B}_X$ such that

$$\mu g(x,y) + \upsilon G(x,y) + \left\langle \left(x_g^*, x_G^*, x_f^* - x_V^* \right), \xi \right\rangle \geqslant \beta. \tag{3.1}$$

Now, we are able to give necessary optimality conditions for the bilevel optimization problem (P). Moreover, under the regularity of Definition 4, we detect Lagrange–Kuhn–Tucker multipliers.

THEOREM 1. *Let $(\overline{x}, \overline{y})$ be a solution of (P). Assume that $\partial^* F$, $\partial^* f$, $\partial^* g$ and $\partial^* G$ are upper semicontinuous at $(\overline{x}, \overline{y})$. Also, suppose that there exists a neighborhood $U \subset X$ of $(\overline{x}, \overline{y})$ such that the functions F, f, g, G are continuous on U and admit bounded convexificators $\partial^* F(\overline{x}, \overline{y})$, $\partial^* f(\overline{x}, \overline{y})$, $\partial^* g(\overline{x}, \overline{y})$ and $\partial^* G(\overline{x}, \overline{y})$ at $(\overline{x}, \overline{y})$.*

Then, there exist scalars λ_1, λ_2, $\gamma \geqslant 0$ and vectors $(\mu_1, \ldots, \mu_{m_1}) \in \mathbb{R}_+^{m_1}$, $(\upsilon_1, \ldots, \upsilon_{m_2}) \in \mathbb{R}_+^{m_2}$ such that

$$
\begin{cases}
\|(\mu, \upsilon, \gamma)\| = 1 \text{ and } \lambda_1 + \lambda_2 = 1, \\
\sum_{i=1}^{m_1} \mu_i g_i(\overline{x}, \overline{y}) = 0 \text{ and } \sum_{j=1}^{m_2} \upsilon_j G_j(\overline{x}, \overline{y}) = 0, \\
(0,0) \in \lambda_1 \operatorname{co} \partial^* F(\overline{x}, \overline{y}) + \lambda_2 \gamma \operatorname{co} \partial^* f(\overline{x}, \overline{y}) + \lambda_2 \sum_{i=1}^{m_1} \mu_i \operatorname{co} \partial^* g_i(\overline{x}, \overline{y}) \\
\quad + \lambda_2 \sum_{j=1}^{m_2} \upsilon_j \operatorname{co} \partial^* G_j(\overline{x}, \overline{y}) - \lambda_2 \gamma (\partial^* V(\overline{x}) \times \{0\})
\end{cases}
$$

where

$$
\begin{cases}
\partial^* V(\overline{x}) = \operatorname{co} \{\partial^* f(.,y)(\overline{x}) : y \in J(\overline{x})\}, \\
J(\overline{x}) = \{y \in \mathbb{R}^{n_2} : g(\overline{x}, y) \leqslant 0 \text{ and } f(\overline{x}, y) = V(\overline{x})\}.
\end{cases}
$$

If in addition to the above assumptions, the problem (P) is regular at $(\overline{x}, \overline{y})$, one has

$$\lambda_1 > 0.$$

Proof. The proof of this theorem consists on several steps.

- Let us prove that there exist scalars λ_1, λ_2, $\gamma \geqslant 0$ and vectors $\mu = (\mu_1, \ldots, \mu_{m_1}) \in \mathbb{R}_+^{m_1}$, $\upsilon = (\upsilon_1, \ldots, \upsilon_{m_2}) \in \mathbb{R}_+^{m_2}$ such that

  $$\|(\mu, \upsilon, \gamma)\| = 1, \ \lambda_1 + \lambda_2 = 1 \text{ and}$$

  $$
  \begin{aligned}
  (0,0) \in \lambda_1 \operatorname{co} \partial^* F(\overline{x}, \overline{y}) &+ \lambda_2 \gamma \operatorname{co} \partial^* f(\overline{x}, \overline{y}) + \lambda_2 \sum_{i=1}^{m_1} \mu_i \operatorname{co} \partial^* g_i(\overline{x}, \overline{y}) \\
  &+ \lambda_2 \sum_{j=1}^{m_2} \upsilon_j \operatorname{co} \partial^* G_j(\overline{x}, \overline{y}) - \lambda_2 \gamma \left(\partial^* V(\overline{x}) \times \{0\}\right);
  \end{aligned}
  $$

where

$$\begin{cases} \partial^* V(\overline{x}) = \text{co}\{\partial^* f(.,y)(\overline{x}) : y \in J(\overline{x})\}, \\ J(\overline{x}) = \{y \in \mathbb{R}^{n_2} : g(\overline{x}, y) \leqslant 0 \text{ and } f(\overline{x}, y) = V(\overline{x})\}. \end{cases}$$

Let $(\overline{x}, \overline{y})$ is an optimal solution to (P). According to Stephane Dempe [9], it is also an optimal solution of (P^*).
Set

$$H(x,y) := (g(x,y), G(x,y), f(x,y) - V(x)) + \mathbb{R}_+^{m_1} \times \mathbb{R}_+^{m_2} \times \mathbb{R}_+,$$

$$\Psi_1(x,y) := F(x,y) - F(\overline{x}, \overline{y}) + \frac{1}{n},\ \Psi_2(x,y) := d(0, H(x,y))$$

and

$$h_n(x,y) := \max(\Psi_1(x,y), \Psi_2(x,y));$$

In this case, $Y_H^* = \mathbb{R}_-^{m_1} \times \mathbb{R}_-^{m_2} \times \mathbb{R}_-$. We have also

$$h_n(\overline{x}, \overline{y}) \leqslant \frac{1}{n} + \inf_{(x,y)\in\mathbb{R}^{n_1}\times\mathbb{R}^{n_2}} h_n(x,y).$$

By using Ekeland's Variational Principle [12], there exists $(x_n, y_n) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ such that

$$\begin{cases} \|(x_n, y_n) - (\overline{x}, \overline{y})\| \leqslant \frac{1}{\sqrt{n}} \\ h_n(x_n, y_n) \leqslant h_n(x,y) + \frac{1}{\sqrt{n}} \|(x,y) - (x_n, y_n)\| \quad \text{for all } (x,y) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}. \end{cases}$$

Hence (x_n, y_n) is a minimum of $h_n(x,y) + \frac{1}{\sqrt{n}} \|(x,y) - (x_n, y_n)\|$ and we get

$$0 \in \text{cl co } \partial^* \left(h_n + \frac{1}{\sqrt{n}} \|. - (x_n, y_n)\| \right)(x_n, y_n).$$

Consequently,

$$0 \in \text{cl co } \partial^* h_n(x_n, y_n) + \frac{1}{\sqrt{n}} \mathbb{B}_{\mathbb{R}^{n_1+n_2}}.$$

In view of Corollary 3, it follows that

$$\partial^* h_n \subset \text{co}\left\{\partial^* \Psi_i : i \in I(x_n, y_n)\right\},$$

where $I(x_n, y_n) := \{i : h_n(x_n, y_n) = \Psi_i(x_n, y_n)\}$.
Consequently, there exists $\lambda_{n,1}, \lambda_{n,2} \in [0,1]$ such that $\lambda_{n,1} + \lambda_{n,2} = 1$ and

$$0 \in \lambda_{n,1} \mathrm{co}\, \partial^* \Psi_1(x_n, y_n) + \lambda_{n,2} \mathrm{co}\, \partial^* \Psi_2(x_n, y_n) + \frac{1}{\sqrt{n}} \mathbb{B}_{\mathbb{R}^{n_1+n_2}} \tag{3.2}$$

where $\lambda_{n,1} = 0$ if $\Psi_1(x_n, y_n) < \Psi_2(x_n, y_n)$, $\lambda_{n,2} = 0$ if $\Psi_2(x_n, y_n) < \Psi_1(x_n, y_n)$, and $0 < \lambda_{n,1} < 1$, $0 < \lambda_{n,2} < 1$ if $\Psi_1(x_n, y_n) = \Psi_2(x_n, y_n)$.
Moreover $\max(\Psi_1(x_n, y_n), \Psi_2(x_n, y_n)) > 0$, otherwise

$$\begin{cases} d(0, H(x_n, y_n)) = 0, \\ F(x_n, y_n) - F(\overline{x}, \overline{y}) + \dfrac{1}{n} = 0. \end{cases}$$

So that $0 \in H(x_n, y_n)$ and $F(x_n, y_n) - F(\overline{x}, \overline{y}) = -\frac{1}{n}$. Since $(\overline{x}, \overline{y})$ is an optimal solution of the problem (P), one has $F(x_n, y_n) - F(\overline{x}, \overline{y}) \geqslant 0$, a contradiction.

From Equation 3.2, and using Corollary 5, there exist $z_n^* = (\mu_n, \upsilon_n, \gamma_n) \in \mathbb{R}_-^{m_1} \times \mathbb{R}_-^{m_2} \times \mathbb{R}_-$ such that $\|z_n^*\| = 1$, $\lambda_{n,1}, \lambda_{n,2} \in [0,1]$, $\lambda_{n,1} + \lambda_{n,2} = 1$ and

$$0 \in \lambda_{n,1} \,\mathrm{co}\, \partial^* F(x_n, y_n) - \lambda_{n,2} \,\mathrm{co}\, \partial^* C_H\left(z_n^*, .\right)(x_n, y_n) + \frac{1}{\sqrt{n}} \mathbb{B}_{\mathbb{R}^{n_1+n_2}}. \tag{3.3}$$

With $\mu_n = \left(\mu_{n,1}, \ldots, \mu_{n,m_1}\right)$, $\upsilon_n = \left(\upsilon_{n,1}, \ldots, \upsilon_{n,m_2}\right)$ and

$$\begin{aligned} \partial^* C_H\left(z_n^*, .\right)(x_n, y_n) = & \sum_{i=1}^{m_1} \mu_{i,n} \partial^* g_i(x_n, y_n) + \sum_{j=1}^{m_2} \upsilon_{j,n} \partial^* G_j(x_n, y_n) \\ & + \gamma_n \partial^* f(x_n, y_n) - \gamma_n \left(\partial^* V(x_n) \times \{0\}\right). \end{aligned}$$

Taking a subsequence if necessary, we can assume that $(\lambda_{n,1}) \to \lambda_1 \in [0,1]$, $(\lambda_{n,2}) \to \lambda_2 \in [0,1]$, $\mu_{i,n} \to \mu_i \leqslant 0$, $\upsilon_{j,n} \to \upsilon_j \leqslant 0$, $\gamma_n \to \gamma \leqslant 0$, $z_n^* \to \widetilde{z}^* = (\mu, \upsilon, \gamma)$ and $\|(\mu, \upsilon, \gamma)\| = 1$ when n tends to $+\infty$.

Then, taking $\left(\mu_i^o, \upsilon_j^o, \gamma^o\right) := -\left(\mu_i, \upsilon_j, \gamma\right)$, one has $\left(\mu_i^o, \upsilon_j^o, \gamma^o\right) \in \mathbb{R}_+^{m_1} \times \mathbb{R}_+^{m_2} \times \mathbb{R}_+$ and

$$\begin{aligned} 0 \in \lambda_1 \,\mathrm{co}\, \partial^* F(\overline{x}, \overline{y}) + \lambda_2 \Bigg[& \sum_{i=1}^{m_1} \mu_i^o \mathrm{co}\, \partial^* g_i(\overline{x}, \overline{y}) + \sum_{j=1}^{m_2} \upsilon_j^o \mathrm{co}\, \partial^* G_j(\overline{x}, \overline{y}) \\ & + \gamma^o \mathrm{co}\, \partial^* f(\overline{x}, \overline{y}) - \gamma^o \left(\partial^* V(\overline{x}) \times \{0\}\right) \Bigg]. \end{aligned}$$

Using Corollary 4, $\operatorname{co}\{\partial^* f\,(.,y)\,(\overline{x}) : y \in J\,(\overline{x})\}$ can be taken as a convexificator of V at $\overline{x}$. We remind the reader that

$$J(\overline{x}) = \left\{ y \in \mathbb{R}^{n_2} : g\,(\overline{x}, y) \leqslant 0 \text{ and } f\,(\overline{x}, y) = V\,(\overline{x}) \right\}.$$

- Let us prove that $\sum_{i=1}^{m_1} \mu_i g_i\,(\overline{x}, \overline{y}) = 0$ and $\sum_{j=1}^{m_2} \upsilon_j G_j\,(\overline{x}, \overline{y}) = 0.$

On the one hand, since $0 \in H\,(\overline{x}, \overline{y})$, we have $C_H\,(\,\widetilde{z}^*, (\overline{x}, \overline{y})) \geqslant 0$. That is,

$$\sum_{i=1}^{m_1} \mu_i g_i\,(\overline{x}, \overline{y}) + \sum_{j=1}^{m_2} \upsilon_j G_j\,(\overline{x}, \overline{y}) + \gamma\,[f\,(\overline{x}, \overline{y}) - V\,(\overline{x})] \geqslant 0. \tag{3.4}$$

On the other hand,

$$g_i\,(\overline{x}, \overline{y}) \leqslant 0,\; G_j\,(\overline{x}, \overline{y}) \leqslant 0 \text{ and } f\,(\overline{x}, \overline{y}) - V(\overline{x}) = 0.$$

Then,

$$\sum_{i=1}^{m_1} \mu_i^o g_i\,(\overline{x}, \overline{y}) + \sum_{j=1}^{m_2} \upsilon_j^o G_j\,(\overline{x}, \overline{y}) + \gamma^o\,[f\,(\overline{x}, \overline{y}) - V\,(\overline{x})] \leqslant 0. \tag{3.5}$$

Combining Equations 3.4 and 3.5, we obtain

$$\sum_{i=1}^{m_1} \mu_i g_i\,(\overline{x}, \overline{y}) + \sum_{j=1}^{m_2} \upsilon_j G_j\,(\overline{x}, \overline{y}) + \gamma\,[f\,(\overline{x}, \overline{y}) - V\,(\overline{x})] = 0.$$

Consequently,

$$\sum_{i=1}^{m_1} \mu_i g_i\,(\overline{x}, \overline{y}) = 0 \text{ and } \sum_{j=1}^{m_2} \upsilon_j G_j\,(\overline{x}, \overline{y}) = 0.$$

- Let us prove that $\lambda_1 > 0$ if (P) is regular at $(\overline{x}, \overline{y})$.

Under the regularity assumption of (P) at $(\overline{x}, \overline{y})$, one can prove that $\lambda_1 > 0$. Indeed, by Equation 3.3 we can choose $x_{1n}^* \in \operatorname{co} \partial^* F\,(x_n, y_n)$, $x_{2n}^* \in \operatorname{co} \partial^* C_H\left(z_n^*, .\right)(x_n, y_n)$ and $x_{3n}^* \in \mathbb{B}_{\mathbb{R}^{n_1+n_2}}$ such that

$$\left(1 - \lambda_{n,1}\right) x_{2n}^* = \lambda_{n,1}\, x_{1n}^* + \frac{1}{\sqrt{n}}\, x_{3n}^*. \tag{3.6}$$

Since (P) is regular at $(\overline{x}, \overline{y})$, there exists $\xi_n \in \delta \mathbb{B}_{\mathbb{R}^{n_1}}$ such that

$$\sum_{i=1}^{m_1} \mu_{i,n} g_i\,(x_n, y_n) + \sum_{j=1}^{m_2} \upsilon_{j,n} G_j\,(x_n, y_n) + \left\langle x_{2n}^*, \xi_n \right\rangle \geqslant \beta. \tag{3.7}$$

Combining Equations 3.6 and 3.7, it yields

$$\sum_{i=1}^{m_1} \mu_{i,n} g_i(x_n, y_n) + \sum_{j=1}^{m_2} \upsilon_{j,n} G_j(x_n, y_n) + \lambda_{n,1} \left\langle x_{1n}^*, \xi_n \right\rangle + \frac{1}{\sqrt{n}} \delta \geqslant \left(1 - \lambda_{n,1}\right) \beta.$$

As $\left\| x_{1n}^* \right\| \leqslant \alpha := \sup_{x^* \in \mathrm{co}\, \partial^* F(\overline{x}, \overline{y})} \|x^*\|$, we have also

$$\sum_{i=1}^{m_1} \mu_{i,n} g_i(x_n, y_n) + \sum_{j=1}^{m_2} \upsilon_{j,n} G_j(x_n, y_n) + \lambda_{n,1} \alpha + \frac{1}{\sqrt{n}} \delta \geqslant \left(1 - \lambda_{n,1}\right) \beta.$$

Letting $n \to +\infty$, we obtain

$$\sum_{i=1}^{m_1} \mu_i g_i(\overline{x}, \overline{y}) + \sum_{j=1}^{m_2} \upsilon_j G_j(x_n, y_n) + \lambda_1 \alpha \geqslant (1 - \lambda_1) \beta.$$

From the feasibility of $(\overline{x}, \overline{y})$, one gets $\lambda_1 \delta \alpha \geqslant (1 - \lambda_1) \beta$. Then, $\lambda_1 \geqslant \frac{\beta}{\alpha\delta + \beta} > 0$. □

Remark 4. A Similar result to Theorem 6 can be obtained by using the notion of approximation introduced for the first time by Jourani and Thibault [18] and revised after by Allali and Amahroq [1].

Remark 5. By a similar argument to that used in [1], one can prove that a convexificator which is upper semicontinuous is an approximation. The converse is true for closed approximations (the argument is similar to that of Proposition 1).

Remark 6. Since the convex hull of a convexificator of a locally Lipschitz function is contained in the Clarke subdifferential, one can deduce another result in terms of the Clarke subdifferential.

Acknowledgement

Thanks are due to Professor Stephane Dempe for helpful remarks and suggestions.

References

1. Allali, K. and Amahroq, T. (1997), Second order approximations and primal and dual necessary optimality conditions, *Optimization*, 3, 229–246.

2. Amahroq, T. and Gadhi, N. (2001), On the regularity condition for vector programming problems, *Journal of Global Optimization*, 21, 435–443.
3. Bard, J.F. (1991), Some properities of the bilevel programming problem, *Journal of optimization theory and applications*, 68, 371–378.
4. Clarke, F.H. (1983), *Optimization and Nonsmooth Analysis*, Wiley-Interscience.
5. Clarke, F.H. (1976), Necessary conditions for a general control problem in calculus of variations and control. in Russel, D. (ed.), *Mathematics Research Center*, Pub. 36, University of Wisconsin, New York academy, pp. 259–278.
6. Chen, Y. and Florian, M. (1992), On the geometry structure of Linear Bilevel Programs: A dual approach, technical Report CRT-867, Centre de Recherche sur les transports, Université de Montreal, Montreal, Quebec, Canada.
7. Craven, B.D., Ralph, D. and Glover, B.M. (1995), Small convex-valued subdifferentials in mathematical programming, *Optimization*, 32, 1–21.
8. Dempe, S. (1992), A necessary and a sufficient optimality condition for bilevel programming problem, *Optimization*, 25, 341–354.
9. Dempe, S. (1997), First-order necessary optimality conditions for general bilevel programming problems, *Journal Of Optimization Theory and Applications*, 95, 735–739.
10. Demyanov, V.F. and Jeyakumar, V. (1997), Hunting for a smaller convex subdifferential, *Journal of Global Optimization*, 10, 305–326.
11. Dien, P.H. (1983), Locally Lipschitzian set-valued maps and general extremal problems with inclusion constraints, *Acta Math Vietnamica*, 1, 109–122.
12. Ekeland, I. (1974), On the variational principle, *Journal of Mathematical Analysis and Application*, 47, 324–353.
13. Huang, H.X. and Pardalos, P.M. (2002), A multivariate partition approach to optimization problems, *Cybernetics and Systems Analysis*, 38 (2), 265–275.
14. Ioffe, A.D. (1989), Approximate subdifferential and applications. III: The metric theory, *Mathematika*, 36, 1–38.
15. Jeyakumar, V. and Luc, D.T. (1998), Approximate Jacobian matrices for continuous maps and C^1-Optimization, *SIAM Journal on Control and Optimization*, 36, 1815–1832.
16. Jeyakumar, V., Luc, D.T. and Schaible, S. (1998), Characterizations of generalized monotone nonsmooth continuous maps using approximate Jacobians, *Journal of Convex Analysis*, 5, 119–132.
17. Jeyakumar, V. and Luc, D.T. (1999), Nonsmooth calculus, minimality, and monotonicity of convexificators, *Journal of Optimization Theory and Applications*, 101, 599–621.
18. Jourani, A. and Thibault, L. (1988), Approximations and metric regularity in mathematical programming in Banach spaces, *Mathematics of Operations Research*, 18, 73–96.
19. Loewen, P.D. (1992), Limits of Frechet normals in nonsmooth analysis, *Optimization and Nonlinear Analysis, Pitman Research Notes Math, Ser*., 244, 178–188.
20. Luderer, B. (1983), Über der Äquivalenz nichtlinearer Optimierungsaufgaben, Technical Report, Technische Universität Karl-Marx-Stadt, Germany.
21. Michel, P. and Penot, J.-P. (1984), Calcul sous-differentiel pour des fonctions Lipschitziennes et non Lipschitziennes, C.R. Acad. Sc. Paris 298.
22. Michel, P. and Penot, J.-P. (1992), A generalized derivative for calm and stable functions, *Differential and Integral Equations*, 5(2), 433–454.
23. Migdalas, A., Pardalos, P.M. and Varbrand, P. (1997), *Multilevel Optimization: Algorithms and Applications,* Kluwer Academic Publishers.
24. Mordukhovich, B.S. and Shao, Y. (1995), On nonconvex subdifferential calculus in Banach spaces, *Journal of Convex Analysis*, 2, 211–228.
25. Outrata, J.V. (1993), On necessary optimality conditions for stackelberg problems, *Journal of Optimiztion Theory and Applications*, 76, 306–320.

26. Outrata, J.V. (1993), Optimality problems with variational inequality constraints, *SIAM Journal on Optimization*, 4, 340–357.
27. Penot, J.P. (1988), On the mean-value theorem, *Optimization*, 19, 147–156.
28. Thibault, L. (1991), On subdifferentials of optimal value functions, *SIAM Journal of Control and Optimization*, 29, 1019–1036.
29. Treiman, J.S. (1995), The linear nonconvex generalized gradient and Lagrange multipliers, *SIAM Journal on Optimization*, 5, 670–680.
30. Wang, S., Wang, Q. and Romano-Rodriguez, S. (1993), Optimality conditions and an algorithm for linear-quadratic bilevel programs, *Optimization*, 4, 521–536.
31. Ye, J.J. and Zhu, D.L. (1995), Optimality conditions for bilevel programming problems, *Optimization*, 33, 9–27.
32. Zhang, R. (1993), Problems of hierarchical optimization in finite dimensions, *SIAM Journal On Optimization*, 4, 521–536.
33. Zowe, J. and Kurcyusz, S. (1979), Regularity and stability for the mathematical programming problem in Banach spaces, *Applied Mathematics and optimization*, 5, 49–62.