Comput Optim Appl (2011) 49: 1–15
DOI 10.1007/s10589-009-9295-9

Direct search algorithm for bilevel programming problems

Ayalew Getachew Mersha · Stephan Dempe

Received: 16 February 2009 / Published online: 30 September 2009

Abstract In this paper, we study the application of a class of direct search methods to bilevel programming with convex lower level problems with strongly stable optimal solutions. In those methods, directions of descent in each iterations are selected within a finite set of directions. To guarantee the existence of such a finite set, we investigate the relation between the aperture of a descent cone at a non stationary point and the vector density of a finite set of directions. It is shown that the direct search method converges to a Clarke stationary point of the bilevel programming problem.

Keywords Bilevel programming · Nonsmooth optimization · Solution algorithm · Direct search algorithm · Clarke stationary solution

1 Introduction

Bilevel programming problems are optimization problems whose feasible set is restricted (in part) to the solution set mapping of another optimization problem. They have a hierarchical (or nested) structure and can be considered as a version of a non-cooperative, two person game which was introduced and investigated by the German economist H. von Stackelberg [20] in 1934. Generally speaking, the decision vector is partitioned among the players: the upper level decision maker and the lower level decision maker in such a way that the upper level decision maker, the leader, controls

Work of the first author was supported by DAAD (Deutscher Akademischer Austausch Dienst) with a scholarship grant.

A.G. Mersha · S. Dempe (✉)
Department of Mathematics and Computer Science, Technical University Bergakademie, Freiberg, Germany
e-mail: dempe@tu-freiberg.de

A.G. Mersha
e-mail: ayalew@math.tu-freiberg.de

one part, say x, and the lower level decision maker, the follower, controls the other part, say y. It should be underscored that each player wants to optimize his respective objective function and, hence, perfect information is assumed. In this paper we consider the following bilevel programming problem:

$$F(x, y) \to \text{“}\min_x\text{”} \tag{1.1}$$

where y solves

$$\begin{aligned} f(x, y) &\to \min_y \\ g(x, y) &\le 0 \\ h(x, y) &= 0 \end{aligned} \tag{1.2}$$

where $F, f : \mathbb{R}^{n+m} \to \mathbb{R}$, $g : \mathbb{R}^{n+m} \to \mathbb{R}^p$, $h : \mathbb{R}^{n+m} \to \mathbb{R}^q$ are sufficiently smooth functions and the lower level problem (1.2) is a convex optimization problem. Define the solution set mapping $\Psi(\cdot)$ by

$$\Psi(x) := \arg\min_y \{f(x, y) : g(x, y) \le 0, h(x, y) = 0\}. \tag{1.3}$$

If the solution of the lower level problem (1.2) corresponding to any parameter x is not unique, the upper level problem (1.1) is not a well defined optimization problem. In this case, the bilevel programming problem may not be solvable. This situation is depicted by examples constructed in [7] and [11, example on p. 121]. The following Lemma substantiates this claim.

Lemma 1.1 [8] *If $\Psi(x)$ is not a singleton for all parameter values x, the leader may not achieve his infimum objective function value.*

To overcome such an unpleasant situation, there are three strategies available for the leader. The first strategy is to replace min with inf in the formulation of problem (1.1) and to define ε-optimal solutions. The second strategy is to allow cooperation between the leader and the follower. This resulted in the so called optimistic or weak bilevel programming problem. The third is a conservative strategy. In this case the leader is forced to bound the damage caused by the follower's “unfavorable” choice. In this paper we will avoid this unpleasant situation by assuming that the solution of the lower level problem (1.2) is uniquely determined for all selections of the leader. First we define regularity conditions for the parametric lower lever problem (1.2). Let

$$\begin{aligned} M &:= \{(x, y) : g(x, y) \le 0, h(x, y) = 0\}, \\ M(X) &:= \{x \in \mathbb{R}^n : \exists y \text{ s.t. } (x, y) \in M\}, \\ M(x) &:= \{y \in \mathbb{R}^m : g(x, y) \le 0, h(x, y) = 0\}, \\ I(x_\circ, y_\circ) &:= \{i : g_i(x_\circ, y_\circ) = 0\}. \end{aligned}$$

Definition 1.2 The lower level problem (1.2) is said to satisfy the *Mangasarian-Fromowitz Constraint Qualification* (MFCQ) at $(x_\circ, y_\circ)$, $y_\circ \in \Psi(x_\circ)$ if

$$\left\{ r \in \mathbb{R}^m \,\middle|\, \begin{array}{l} r^\top \nabla_y g_i(x_\circ, y_\circ) < 0, \forall i \in I(x_\circ, y_\circ), \\ r^\top \nabla_y h_j(x_\circ, y_\circ) = 0, \; j = 1, 2, \ldots q \end{array} \right\} \ne \emptyset$$

and the gradients $\{\nabla_y h_j(x_\circ, y_\circ) : j = 1, \dots, q\}$ are linearly independent.

The Lagrangian function for the lower level problem is given by

$$L(x, y, \lambda, \mu) := f(x, y) + \lambda^\top g(x, y) + \mu^\top h(x, y).$$

Consider the set of Lagrange multipliers

$$\Lambda(x, y) := \{(\lambda, \mu) : \lambda \geq 0, \lambda^\top g(x, y) = 0, \nabla_y L(x, y, \lambda, \mu) = 0\}.$$

Let

$$J(\lambda) := \{j : \lambda_j > 0\}.$$

Definition 1.3 The lower level problem (1.2) is said to satisfy a *strong sufficient optimality condition of second order* (SSOC) at a point $(x_\circ, y_\circ)$ if for each $(\lambda, \mu) \in \Lambda(x_\circ, y_\circ)$ and for every nonzero element of the set

$$\left\{ r \in \mathbb{R}^m \,\middle|\, \begin{array}{l} r^\top \nabla_y g_i(x_\circ, y_\circ) = 0, \forall i \in J(\lambda), \\ r^\top \nabla_y h_j(x_\circ, y_\circ) = 0, \, j = 1, 2, \dots, q \end{array} \right\}$$

we have

$$r^T \nabla^2_{yy} L(x_\circ, y_\circ, \lambda, \mu) r > 0.$$

Definition 1.4 The *constant rank constraint qualification* (CRCQ) is valid for problem (1.2) at a point $(x_\circ, y_\circ)$ if there exists an open neighborhood $W_\varepsilon(x_\circ, y_\circ)$, $\varepsilon > 0$ of $(x_\circ, y_\circ)$ such that for each subsets $I \subseteq I(x_\circ, y_\circ)$, $J \subseteq \{1, \dots, q\}$ the family of gradient vectors $\{\nabla_y g_i(x, y) : i \in I\} \cup \{\nabla_y h_j(x, y) : j \in J\}$ has the same rank for all $(x, y) \in W_\varepsilon(x_\circ, y_\circ)$.

Let the following assumption be satisfied:

Assumption 1 *The set M is nonempty and compact.*

2 Piecewise continuously differentiable functions

Definition 2.1 A function $y : \mathbb{R}^n \to \mathbb{R}^m$ is a *piecewise continuously differentiable* (or PC^1) function at $x_\circ$ if it is continuous and there exists an open neighborhood V of $x_\circ$ and a finite number of continuously differentiable functions $y^i : V \to \mathbb{R}^m$, $i = 1, \dots, b$ such that $y(x) \in \{y^1(x), y^2(x), \dots, y^b(x)\}$ for all $x \in V$. The function y is a PC^1 function on some open set O provided it is a PC^1 function at every point $x_\circ \in O$.

Theorem 2.2 [16] *Let the lower level problem* (1.2) *at $x = x_\circ$ be convex satisfying* (*MFCQ*), (*SSOC*) *and* (*CRCQ*) *at a stationary solution $y_\circ$. Then, the locally uniquely determined function $y(x) \in \Psi(x)$, and hence $F(x, y(x))$ is a PC^1 function.*

It has been shown in [12] that PC^1 functions are locally Lipschitz continuous. The existence of the directional derivative

$$y'(x;d) := \lim_{t\downarrow 0} t^{-1}[y(x+td) - y(x)]$$

and, hence, directional differentiability of the function $F(x, y(x))$ is an implication of a result in [14]. This is restated as

Theorem 2.3 [11, Theorem 4.8] *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a PC^1 function. Then f is directionally differentiable.*

Under the assumptions (MFCQ), (CRCQ), and (SSOC) the bilevel programming problem (1.1) can be replaced by a single level Lipschitz optimization problem:

$$\mathcal{F}(x) := F(x, y(x)) \to \min_x . \tag{2.1}$$

The following theorem gives necessary optimality conditions. It shows the nonexistence of a feasible descent direction at a local optimal solution of the bilevel programming problem.

Theorem 2.4 [10] *Let $(x_\circ, y_\circ)$, with $y_\circ \in \Psi(x_\circ)$, be a local optimal solution of the bilevel programming problem (1.1), (1.2) and assume that the lower level problem (1.2) is a convex parametric optimization problem satisfying (MFCQ), (SSOC) and (CRCQ) at $(x_\circ, y_\circ)$. Then the following problem has a non negative optimal objective function value:*

$$\alpha \to \min_{\alpha,d} \tag{2.2a}$$

$$\nabla_x F(x_\circ, y_\circ)d + \nabla_y F(x_\circ, y_\circ)y'(x_\circ; d) \le \alpha \tag{2.2b}$$

$$\|d\| \le 1. \tag{2.2c}$$

Note that the convexity assumptions can not be relaxed because without convexity, even under the assumptions made in the theorem, the lower level optimal solution may not be unique. In the direct search algorithm, a Clarke stationary solution of problem (1.1), (1.2) will be computed. To determine the Clarke generalized differential of the function $\mathcal{F}(x) := F(x, y(x))$ we need the index set of essentially active selection functions $I_y^s(x)$. Let the PC^1-function $y(x)$ be a continuous selection of the continuously differentiable functions $y_i(x)$, $i = 1, \ldots, b$. Then,

$$\mathcal{K}(i) := \{x : y(x) = y_i(x)\}$$

is the set of all points x, where the PC^1-function coincides with one of its selection functions, and

$$I_y^s(x) := \{i : y(x) = y_i(x), x \in \text{cl int}\mathcal{K}(i)\} \subseteq \{1, 2, \ldots, b\}$$

is the index set of essentially active selection functions. The tangent (or Bouligand) cone to the set $\mathcal{K}(i)$ is defined as

$$T_{\mathcal{K}(i)}(x) := \left\{ d : \exists x_k \to x, x_k \in \mathcal{K}(i)\ \forall k, t_k \searrow 0 \text{ such that } \frac{x_k - x}{t_k} \to d \right\}.$$

Once the set of essentially active selection function is known, it has been shown in Scholtes [17, Proposition A.4.1], that the generalized Jacobian of $y(\cdot)$ is given by

$$\partial^\circ y(x) = \mathrm{conv}\{\nabla y_i(x) : i \in I_y^s(x)\}.$$

The generalized directional derivative (in the sense of Clarke) of the solution function $y(\cdot)$ is given by

$$y^\circ(x; d) = \sup\{\langle d, z\rangle : z \in \partial^\circ y(x)\}.$$

Definition 2.5 A point $(x_\circ, y_\circ)$, $y_\circ \in \Psi(x_\circ)$ is said to be a *Clarke stationary point* for the bilevel programming problem (1.1), (1.2) if $\forall d$, $\exists i \in I_y^s(x_\circ)$ such that $d \in T_{\mathcal{K}(i)}(x_\circ)$, $\nabla_x F(x_\circ, y_\circ)d + \nabla_y F(x_\circ, y_\circ)\nabla y_i(x_\circ)d \geq 0$.

In other words, the point $(x_\circ, y_\circ)$ is Clarke stationary if $\mathcal{F}^\circ(x_\circ; d) \geq 0$ for all directions d [9]. The necessary optimality condition is given by:

Theorem 2.6 *Let the lower level problem* (1.2) *satisfy* (*MFCQ*), (*SSOC*) *and* (*CRCQ*). *If a point* $(x_\circ, y_\circ)$ *with* $y_\circ \in \Psi(x_\circ)$ *is a local minimum of the bilevel programming problem* (1.1), (1.2), *then it is a Clarke stationary point for the bilevel programming problem.*

Proof Let $(x_\circ, y_\circ)$ be local optimal solution of (1.1), (1.2). From Theorem 2.4 we have

$$\nabla_x F(x_\circ, y_\circ)d + \nabla_y F(x_\circ, y_\circ)y'(x_\circ; d) \geq 0, \quad \forall d.$$

Then there is $i \in I_y^s(x_\circ)$ with $d \in T_{\mathcal{K}(i)}(x_\circ)$, $y'(x_\circ; d) = \nabla y_i(x_\circ)d$. From Theorem 2.3 we derive

$$\nabla_x F(x_\circ, y_\circ)d + \nabla_y F(x_\circ, y_\circ)y'(x_\circ; d) = \nabla_x F(x_\circ, y_\circ)d + \nabla_y F(x_\circ, y_\circ)\nabla y_i(x_\circ)d \geq 0$$

for $i \in I_y^s(x_\circ)$, $d \in T_{\mathcal{K}(i)}(x_\circ)$. Hence the theorem. □

The direct search algorithm as formulated e.g. in [13] can be applied to problems of minimizing smooth functions. Our aim in this work is to extend it to nonsmooth optimization of a very special type, i.e. to apply it to the class of bilevel programming problem with convex lower level problems having strongly stable optimal solutions.

3 Direct search methods

Direct search methods are iterative optimization methods which do not require the computation or any approximation of gradients. The directions used to update the iterations are selected from a special finite set of directions. These methods are popular in industry. They were used to solve difficult problems that arise from industry and engineering because they can easily be applied to almost any optimization problem, including those with nonsmooth objective functions. It should be noted that the term direct search method is broad. There exist many classes of direct search methods in the literature. These classes include pattern search methods [18, 19]. Convergence proofs for such algorithms can be found in [4, 5, 18, 19] and in the review [13]. The reader is also referred to the references therein for the details of the development of many versions of algorithms and thorough theoretical analysis of their convergence proofs. In [1] Abramson and in [2] Abramson and Audet proposed the second-order convergence analysis for generalized pattern search (GPS). In [13, p. 443], it was reported that there were no theoretical guarantee that some versions of pattern search methods work for a nonsmooth optimization problems. To demonstrate this problem, a simple example is constructed by taking a variant of the Dennis-Woods two dimensional function and apply compass search method to get a local minimizer. The reason for such a failure is that pattern search algorithms explore the space of variables using always the same finite set of directions. In 2006, Audet and Dennis [6] proposed a new class of pattern search methods called the mesh adaptive direct search (MADS) which not only overcomes the limitation of exploring through a finite number of directions but also second-order convergence is given. They established that if the objective function is locally Lipschitz around the limit point, then the limit point is Clarke stationary point under some mild assumptions (see [6, Theorem 3.13]). Though the method used in this paper and the one reported in [15] are motivated by the investigations in pattern search methods, we do not consider pattern search methods or other frame, mesh or grid based methods which are based on a simple decrease of the objective function. Our investigation focuses on the class of direct search methods that are based on a sufficient decrease and hence we do not require neither the step update parameter to be an integer power of some rational number nor the set of directions satisfy special conditions such as reported in [19, Sect. 2] and [3] for the pattern search methods. The term direct search method hereafter should be understood in this sense.

Direct search methods for minimizing a function $\theta(x)$ can be summarized in the following way. Consider the k-th iterate x_k. The next iterate $x_{k+1} \neq x_k$ is produced if there is a scalar $t_k > 0$ and a direction $d_k \in D$ such that $\theta(x_k + t_k d_k) < \theta(x_k)$. In this case we write $x_{k+1} := x_k + t_k d_k$. To realize this strategy, one has to define the set D of search directions. If the mentioned descent step cannot be realized for all directions in D the step size parameter t_k is first reduced.

The definition and characterization of the aforementioned set of directions, D, is the key to the convergence of direct search methods. The existence of such a finite set of search directions at a nonstationary point within a certain compact set will be shown next.

Assume $\theta(\cdot)$ is a locally Lipschitz continuous function near a given point x.

Definition 3.1 [15] A vector d is said to be *descent direction* of $\theta(\cdot)$ at x if $\theta'(x; d) < 0$. A set $D \subset \mathbb{R}^n$ is a *descent set* of $\theta(\cdot)$ in some set $\mho$ if for every $x \in \mho$ there is $d \in D$ such that $\theta'(x; d) < 0$ provided that $\theta(\cdot)$ is directionally differentiable in the direction d.

The next theorem shows the existence of a finite set of descent directions at a nonstationary point.

Theorem 3.2 [15] *Let $\theta(\cdot)$ be a locally Lipschitz function near $x \in \mho$, where $\mho \subset \mathbb{R}^n$ is a compact set that does not contain stationary points of $\theta(\cdot)$. Then there exists a finite set D of vectors and a positive number α such that*

$$\min_{d \in D} \theta'(x; d) \leq -\alpha \tag{3.1}$$

for every $x \in \mho$ provided that $\theta(\cdot)$ is directionally differentiable in the direction d.

The classical directional derivative can be replaced by the generalized directional derivative for lower semicontinuous functions.

4 Application of direct search methods to bilevel programming problems

Let $(x_\circ, y_\circ)$ be an arbitrary point and assume that the set

$$X := \{x : F(x, y(x)) \leq F(x_\circ, y_\circ), y(x) \in \Psi(x)\}$$

is compact and nonempty. Let $\mathfrak{C}$ be the set of Clarke stationary points of the (2.1). Let δ be a sufficiently small positive number. Set

$$S_{\mathfrak{C}} := X \setminus \bigcup_{x \in \mathfrak{C}} B(x, \delta)$$

where $B(x, \delta)$ is an open neighborhood of x with radius $\delta > 0$. Note that the generalized directional derivative of the function $\mathcal{F}$ at some point x is equal to

$$\mathcal{F}^\circ(x; d) = \nabla_x F(x, y)d + \nabla_y F(x, y) y^\circ(x_\circ; d)$$

provided that the function $F(\cdot)$ is continuously differentiable and the function $y(\cdot)$ is locally Lipschitz continuous. Then, the set $S_{\mathfrak{C}}$ is closed.

Theorem 4.1 *Let the lower level problem be convex satisfying* (*MFCQ*), (*SSOC*) *and* (*CRCQ*). *Then there exists a finite set D of vectors and a positive number $\alpha > 0$ such that*

$$\min_{d \in D} \mathcal{F}^\circ(x; d) \leq -\alpha \tag{4.1}$$

for every $x \in S_{\mathfrak{C}}$.

Proof Suppose this is not the case. Then for every finite set of vectors D and any positive number α there is $x \in S_{\mathfrak{C}}$ such that

$$\min_{d\in D} \mathcal{F}^\circ(x; d) > -\alpha.$$

Take an arbitrary sequence $\{\alpha_k\}$ of positive numbers converging to zero and a finite set $D_p = \{d_1, d_2, \dots, d_j\}_{j=1}^p$. The definition of this set will be made more precise later in the proof. For the moment we need only finiteness.

Since (4.1) is not satisfied, for every k there exists $x_k \in S_{\mathfrak{C}}$ such that

$$\min_{d\in D} \mathcal{F}^\circ(x_k; d) > -\alpha_k. \tag{4.2}$$

Then, since the generalized derivative of PC^1-functions is the convex hull of finitely many points, for all i and some essentially active selection function $y_{s_k}(\cdot)$, $s_k \in I_y^s(x_k)$, the following inequality is valid:

$$\begin{aligned} \mathcal{F}^\circ(x_k, d_i) &= \nabla_x F(x_k, y(x_k))d_i + \nabla_y F(x_k, y(x_k))y^\circ(x_k; d_i) \\ &= \nabla_x F(x_k, y(x_k))d_i + \nabla_y F(x_k, y(x_k))\nabla y_{s_k}(x_k)d_i \\ &> -\alpha_k. \end{aligned} \tag{4.3}$$

Due to compactness of the set X, $\{x_k\}$ is bounded and, hence, has an accumulation point, say $x_\circ$. For the sake of simplicity assume without loss of generality that $x_k \to x_\circ$. Since there exist only a finite number of selection functions, for a fixed $d_i \in D$, we can assume that $s = s_k$ with $s_k \in I_y^s(x_k)$ is also fixed (otherwise we take an infinite subsequence). Since the graph of the tangent cone mapping is closed, $d_i \in T_{\mathcal{K}(s_k)}(x_k)$ for all k and fixed i and $x_k \to x_\circ$ imply $d_i \in T_{\mathcal{K}(s)}(x_\circ)$. Now, from (4.2) and (4.3) we have

$$\nabla_x F(x_k, y(x_k))d_i + \nabla_y F(x_k, y(x_k))\nabla y_s(x_k)d_i > -\alpha_k. \tag{4.4}$$

Taking the limit $k \to \infty$ in (4.4), we derive

$$\nabla_x F(x_\circ, y(x_\circ))d_i + \nabla_y F(x_\circ, y(x_\circ))\nabla y_s(x_\circ)d_i \ge 0, \quad \forall d_i \in T_{\mathcal{K}(s)}(x_\circ).$$

Now choose a finite set of vectors $D = \{d_1, d_2, \dots, d_k\}$ such that

$$\forall T_{\mathcal{K}(i)}(x_\circ)\ \exists\{d_{i_1}, d_{i_2}, \dots, d_{i_p}\} \subseteq D$$

satisfying

$$T_{\mathcal{K}(i)}(x_\circ) \subseteq \text{cone}\,\{d_{i_1}, d_{i_2}, \dots, d_{i_p}\},$$

where cone M denotes the conical hull of the set M. Then, $\forall d \in \mathbb{R}^n\ \exists i : d \in T_{\mathcal{K}(i)}(x_\circ)$ and there are nonnegative numbers $\beta_j \ge 0$ such that

$$\begin{aligned} &\nabla_x F(x_\circ, y(x_\circ))d + \nabla_y F(x_\circ, y(x_\circ))\nabla y_s(x_\circ)d \\ &\quad = \sum_{j=1}^p \beta_j(\nabla_x F(x_\circ, y(x_\circ))d_{i_j} + \nabla_y F(x_\circ, y(x_\circ))\nabla y_s(x_\circ)d_{i_j}) \ge 0. \end{aligned} \tag{4.5}$$

This implies $x_\circ$ is Clarke stationary, which contradicts $x_\circ \in S_{\mathfrak{C}}$. Hence, the theorem is correct. □

The *cosine measure* of D, denoted by $\rho(D)$, is used to verify if a given finite set can be used as descent set. It is defined as

$$\rho(D) = \inf_{\substack{u\neq 0\\ u\in\mathbb{R}^n}} \sup_{\substack{v\neq 0\\ v\in D}} \frac{u^\top v}{\|u\|\|v\|}. \tag{4.6}$$

A finite set $D \subset \mathbb{R}^n$ is called *generating* if its conical hull equals $\mathbb{R}^n$. If D is a generating set then $\rho(D) > 0$. The cosine measure can be used to estimate the quality of search directions: it is a measure for the "distance" of D to the steepest descent direction. We impose a lower bound on the cosine measure of a set of search directions in order to "protect" the search directions from being "too close" to the orthogonal to $\nabla f(x_k)$.

Next, define the descent cone and the aperture of a convex cone.

Definition 4.2 Let the lower level problem satisfy (MFCQ), (CRCQ), and (SSOC). A *descent cone* of $F(\cdot, y(\cdot)) = \mathcal{F}(\cdot)$ at a point x is defined as

$$C(x) := \{v \in \mathbb{R}^n : \mathcal{F}^\circ(x; v) < 0\}.$$

Definition 4.3 [15] Let C be a convex cone that does not contain zero vectors, and let $\overline{C}$ be the closure of C. The *aperture of the cone* C, denoted by $\psi(C)$, is defined as

$$\psi(C) := \arccos\left(\min_{\substack{w\in\overline{C}\\ w\neq 0}} \sup_{\substack{z\in\mathbb{R}^n\setminus\overline{C}\\ z\neq 0}} \frac{w^\top z}{\|w\|\|z\|}\right).$$

The aperture of a cone varies between the two extremes 0 and $\frac{\pi}{2}$. The aperture of the descent cone $C(x)$ can be used to show that this cone intersects a finite generating cone, provided that some assumptions are satisfied.

Lemma 4.4 [15] *Let D be a finite generating set with positive vector density $\rho(D)$ and let C be a given open cone with an aperture $\psi(C) > 0$. If $\rho(D) > \cos(\psi(C))$ then $D \cap C \neq \emptyset$.*

Definition 4.5 Let the lower level problem satisfy (MFCQ), (CRCQ) and (SSOC). The *descent aperture* of $\mathcal{F}(\cdot)$ of some set S is defined as the smallest aperture of all descent cones of $\mathcal{F}(.)$ on that set:

$$\phi(\mathcal{F}, S) := \inf_{x\in S} \psi(C(x))$$

where $C(x)$ is a convex cone contained in a descent cone $\{d : \mathcal{F}^\circ(x; d) < 0\}$.

Now we state one of the important results that helps us to prove our main result.

Theorem 4.6 *Let the lower level problem satisfy (MFCQ), (SSOC), (CRCQ) at* $(x_\circ, y_\circ)$, *where* $y_\circ \in \Psi(x_\circ)$. *Assume that the set* $X = \{x : \mathcal{F}(x) \leq \mathcal{F}(x_\circ)\}$ *is compact. Then* $\phi(\mathcal{F}, S_{\mathfrak{C}}) > 0$.

Proof If (MFCQ), (SSOC) and (CRCQ) are assumed to be satisfied for the convex parametric lower level problem, then $y(\cdot)$ is a piecewise continuously differentiable function and hence $\mathcal{F}(\cdot)$ is also a piecewise continuously differentiable function [11]. Therefore $\mathcal{F}(\cdot)$ is locally Lipschitz. By Theorem 4.1 there exists a finite set of nonzero vectors D and a real number $\alpha > 0$ such that

$$\min_{d \in D} \mathcal{F}^\circ(x; d) < -\alpha$$

for all $x \in S_{\mathfrak{C}}$.

Take any $\hat{x}$ from $S_{\mathfrak{C}}$ and select a vector $r \in D$ such that $\mathcal{F}^\circ(\hat{x}, r) \leq -\alpha$. Assume that

$$0 < \beta_{\min} := \min_{d \in D} \|d\| \leq \max_{d \in D} \|d\| =: \beta_{\max} < +\infty.$$

Set $\delta := \min(\beta_{\min}, \frac{\alpha}{2\mathcal{L}})$ where $\mathcal{L}$ is the Lipschitz constant of $\mathcal{F}(x)$ on $S_{\mathfrak{C}}$. Now, the Clarke directional derivative

$$\mathcal{F}^\circ(\hat{x}; z) \leq \mathcal{F}^\circ(\hat{x}; r) + \mathcal{F}^\circ(\hat{x}; z - r) \leq -\alpha + \mathcal{L}\|z - r\| \leq -\alpha + \mathcal{L}\delta \leq -\frac{\alpha}{2}$$

$\forall z \in B_\delta(r) := \{z \in D : \|r - z\| < \delta\}$.

Assume without loss of generality $\beta_{\min} \geq \frac{\alpha}{2\mathcal{L}}$. Construct the cone

$$C(r, \delta) := \{tz : t > 0, z \in B_\delta(r)\}.$$

Let $\hat{d} \in C(r, \delta)$. This implies $\hat{d} = tz$, $z \in B_\delta(r)$ for some $t > 0$.

$$\mathcal{F}^\circ(\hat{x}; \hat{d}) = \mathcal{F}^\circ(\hat{x}; tz) = t\mathcal{F}^\circ(\hat{x}; z) < -t\alpha/2$$

which implies that $\hat{d} \in C(\hat{x})$ for some convex cone $C(\cdot)$ contained in a descent cone of $\mathcal{F}$. Therefore, $C(r, \delta) \subseteq C(\hat{x})$. We have

$$\inf_{\substack{u \in C(r,\delta) \\ u \neq 0}} \sup_{\substack{v \in \mathbb{R}^n \setminus \bar{C}(r,\delta) \\ v \neq 0}} \frac{u^\top v}{\|u\|\|v\|} \geq \inf_{\substack{u \in C(\hat{x}) \\ u \neq 0}} \sup_{\substack{v \in \mathbb{R}^n \setminus \bar{C}(\hat{x}) \\ v \neq 0}} \frac{u^\top v}{\|u\|\|v\|}.$$

This implies

$$\arccos\left(\inf_{\substack{u \in C(r,\delta) \\ u \neq 0}} \sup_{\substack{v \in \mathbb{R}^n \setminus \bar{C}(r,\delta) \\ v \neq 0}} \frac{u^\top v}{\|u\|\|v\|}\right) \leq \arccos\left(\inf_{\substack{u \in C(\hat{x}) \\ u \neq 0}} \sup_{\substack{v \in \mathbb{R}^n \setminus \bar{C}(\hat{x}) \\ v \neq 0}} \frac{u^\top v}{\|u\|\|v\|}\right).$$

Hence, $\psi(C(\hat{x})) \geq \psi(C(r, \delta))$.

On the other hand observe that

$$\beta_{\min} \leq \|d\| \leq \beta_{\max} \quad \forall d \in D \quad \Rightarrow \quad \beta_{\min} \leq \|\hat{d}\| \leq \beta_{\max}$$

$$\Rightarrow \quad \frac{1}{\beta_{\max}} \leq \frac{1}{\|\hat{d}\|} \leq \frac{1}{\beta_{\min}}$$
$$\Rightarrow \quad \frac{\alpha}{2\mathcal{L}\beta_{\max}} \leq \frac{\alpha}{2\mathcal{L}\|\hat{d}\|} \leq \frac{\alpha}{2\mathcal{L}\beta_{\min}}$$

Since $\delta = \frac{\alpha}{2\mathcal{L}}$ or $\frac{\delta}{\|r\|} = \frac{\alpha}{2\mathcal{L}\|r\|}$, the following conclusion is true:

$$\frac{\alpha}{2\mathcal{L}\beta_{\max}} \leq \frac{\delta}{\|r\|} \leq \frac{\alpha}{2\mathcal{L}\beta_{\min}} \quad \Longrightarrow \quad 0 < \arcsin\left(\frac{\alpha}{2\mathcal{L}\beta_{\max}}\right) \leq \arcsin\left(\frac{\delta}{\|r\|}\right).$$

Consider now

$$\begin{aligned} \psi(C(r,\delta)) &= \arccos \inf_{\substack{u \in C(r,\delta) \\ u \neq 0}} \sup_{\substack{v \in \mathbb{R}^n \setminus \bar{C}(r,\delta) \\ v \neq 0}} \frac{u^\top v}{\|u\|\|v\|} \\ &= \arccos \inf_{\substack{z \in B_\delta(r) \\ u \neq 0}} \sup_{\substack{v \in \mathbb{R}^n \setminus \bar{C}(r,\delta) \\ v \neq 0}} \frac{tz^\top v}{\|tz\|\|v\|}, \quad t > 0 \\ &\geq \arccos\left(\frac{\|z\|}{\|r\|}\right), \quad \forall z \in B_\delta(r). \end{aligned} \tag{4.7}$$

This implies $\psi(C(r,\delta)) \geq \arcsin(\frac{\delta}{\|r\|})$. Clearly this inequality holds for all $\hat{x}$ from $S_{\mathfrak{C}}$. Hence, it follows that

$$\phi(\mathcal{F}, S_{\mathfrak{C}}) \geq \arcsin \frac{\alpha}{2\mathcal{L}\beta_{\max}} > 0.$$

□

An important theorem for the characterization of the finite set D follows.

Theorem 4.7 *Let $X = \{x : \mathcal{F}(x) \leq \mathcal{F}(x_\circ)\}$ be compact. Assume that the lower level problem satisfies (MFCQ), (SSOC) and (CRCQ). Let $\mathfrak{C}$ be the set of stationary points of the bilevel programming problem. Then any set with vector density $\rho(D) > cos(\phi(\mathcal{F}, S_{\mathfrak{C}}))$ is a descent set.*

Proof Using similar implications as in the proof of Theorem 4.6, $\mathcal{F}(\cdot)$ is a locally Lipschitz continuous function. Then by Theorem 4.6 we have $\phi(\mathcal{F}, S_{\mathfrak{C}}) > 0$. By Lemma 4.4, if D is finite set of vectors with $\rho(D) > \cos(\phi(\mathcal{F}, S_{\mathfrak{C}}))$, then we have $D \cap C(x)$ is non empty for every $x \in S_{\mathfrak{C}}$ where $C(x)$ is a convex cone contained in a descent cone of $\mathcal{F}(\cdot)$ at x. That means there is $d \in D$ such that $\mathcal{F}^\circ(x; d) < 0$ which implies D is a descent set. □

The following algorithm is a prototype realization of the proposed direct search algorithm. Before we give the convergence proof of this algorithm we state two intermediate results.

Lemma 4.8 *Let the lower level problem be convex satisfying* (*MFCQ*), (*SSOC*) *and* (*CRCQ*) *at all* x. *Let* D *be any finite generating set of vectors such that the vector density* $\rho(D) > \cos(\phi(\mathcal{F}, S_{\mathfrak{C}}))$ *and* $\|d\| \in [\beta_{\min}, \beta_{\max}]$, $\forall d \in D$. *Then for sufficiently small* $\delta > 0$ *and* t^* *there exists* $d \in D$ *such that*

$$\mathcal{F}(x + td) - \mathcal{F}(x) \leq -\delta t$$

for all $t \in (0, t^*)$ *and for all* $x \in S_{\mathfrak{C}}$.

Proof Since the assumptions (MFCQ), (SSOC), and (CRCQ) are satisfied for the lower level problem (1.2) at all x, the function $\mathcal{F}(\cdot)$ is piecewise continuously differentiable, hence directionally differentiable. Since $x \in S_{\mathfrak{C}}$, it follows that x is not a local solution of problem (2.1). Thus, there exists a direction d such that $\mathcal{F}^\circ(x; d) < -2\delta$ for a sufficiently small $\delta > 0$. Hence, $\mathcal{F}'(x; d) < -2\delta$, too. Then

$$\frac{\mathcal{F}(x + td) - \mathcal{F}(x)}{t} < -\delta, \quad \forall t \in (0, t^*]$$

and some $t^* > 0$. It remains to show that the set D contains a direction d with the desired property. By Theorem 4.7, D is a descent set. From Lemma 4.4 we see that D contains d with $\mathcal{F}^\circ(x; d) < -\delta$, $\forall x \in S_{\mathfrak{C}}$. □

Lemma 4.9 *Let* $X = \{x : \mathcal{F}(x) \leq \mathcal{F}(x_\circ)\}$ *be compact. Let* $\{x_k\}$ *be a sequence of iterates produced by Algorithm 1. Then* $\lim_{k\to\infty} t_k = 0$.

Proof Let $\{t_k\}$ be a sequence of step size parameter. Suppose there exist a subsequence $\{t_{k_i}\}_{i=0}^{\infty}$ with $\lim_{i\to\infty} t_{k_i} = \eta > 0$. Then, either $\lim_{k\to\infty} t_k = \eta$ or the limit does not exist. Observe that t_{k_i} is the step size in a successful iteration. Then

$$\mathcal{F}(x_{k_i} + t_{k_i} d_{k_i}) < \mathcal{F}(x_{k_i}) - \omega(t_{k_i}), \quad i = 1, 2, \ldots$$

Algorithm 1 Direct search algorithm for bilevel programming: the first version

Initialization: Select an initial point $x_\circ$, $\rho_\circ > \phi(\mathcal{F}, S_{\mathfrak{C}})$, and take a finite set D with vector density $\rho_\circ$. $k := 0$. Consider constants $c_1, c_2, t_\circ, \beta_{\max}, \beta_{\min}$ with $0 < c_1 < 1 < c_2$, $t_\circ > 0$ sufficiently small and $\beta_{\max} > \beta_{\min}$. Let $\omega(t)$ be a continuous function such that $\lim_{t\to 0} \frac{\omega(t)}{t} = 0$ and $\omega(t) > 0$ if $t > 0$.

Step 1: For the current step size t_k search for $d_k \in D$ such that $\mathcal{F}(x_k + t_k d_k) < \mathcal{F}(x_k) - \omega(t_k)$. If there exists such $d_k \in D$ then go to Step 2 otherwise go to Step 3.

Step 2: Successful step. Do the following:
set $x_{k+1} = x_k + t_k d_k$
set $t_{k+1} = c_{2_k} t_k$, $c_{2_k} \in [1, c_2)$ and go to Step 1.

Step 3: Unsuccessful step. Do the following:
Set $x_{k+1} = x_k$
Set $t_{k+1} = c_{1_k} t_k$, $c_{1_k} \in (c_1, 1)$ and go to step 4.

Step 4: Stopping condition. If $t_{k+1} < t_\circ$ then terminate.
Else put $k := k + 1$ go to Step 1.

with $\omega(t_{k_i}) > 0$. Continuity of $\omega(\cdot)$ and $\omega(t) > 0$ for $t > 0$ implies that we have $\lim_{i\to\infty} \mathcal{F}(x_{k_i}) = -\infty$ which contradicts the boundedness from below of $\mathcal{F}$. Hence, this infinite subsequence can not exist and hence $\lim_{i\to\infty} t_k = 0$. □

Remark 4.10 Observe that a sufficient condition for $\mathcal{F}(\cdot)$ to be bounded from below is that the set $X = \{x : \mathcal{F}(x) \le \mathcal{F}(x_\circ)\}$ is compact for some point $x_\circ$.

Lemma 4.11 *Let $X = \{x : \mathcal{F}(x) \le \mathcal{F}(x_\circ)\}$ be compact and let the lower level problem satisfy* (*MFCQ*), (*SSOC*), *and* (*CRCQ*). *Let $\{x_k\}$ be a sequence of iterates produced by Algorithm 1. Then, there is a subsequence $\{x_{k_i}\}$ of the sequence $\{x_k\}$ that converges to an element of $\mathfrak{C}$.*

Proof By Lemma 4.9 we have $\lim_{i\to\infty} t_{k_i} = 0$. This implies there are infinitely many unsuccessful iterates. Let δ be an arbitrary small positive number and define

$$S_{\mathfrak{C}}(\delta) := X \setminus \bigcup_{x\in\mathfrak{C}} B(x,\delta).$$

Since $x_k \in X$, it follows either $x_k \in S_{\mathfrak{C}}(\delta)$ or $x_k \in \bigcup_{x\in\mathfrak{C}} B(x,\delta)$. Let $x_k \in S_{\mathfrak{C}}(\delta)$. Due to $\rho(D) > \cos(\phi(\mathcal{F}, S_{\mathfrak{C}}))$ and using Lemma 4.8 we obtain that, for such x_k and for all t^* and σ there exists $d \in D$ such that $\mathcal{F}(x_k + td) - \mathcal{F}(x_k) < -\sigma t, \forall t \in (0, t^*]$. The number of such iterates x_k must be finite, since otherwise we have $\mathcal{F}(x_k) \to -\infty$ which is a contradiction to the assumption that $\mathcal{F}(\cdot)$ is bounded below. Hence there must exist $k_\circ$ such that

$$x_k \in \bigcup_{x\in\mathfrak{C}} B(x,\delta), \quad \forall k \ge k_\circ.$$

Since $\delta > 0$ is arbitrarily chosen, this implies that all accumulation points of $\{x_k\}$ belong to $\mathfrak{C}$. By boundedness of X there exists at least one accumulation point. □

The main difficulty in Algorithm 1 is to find $\phi(\mathcal{F}, S_{\mathfrak{C}})$. Clearly for an arbitrary set it is difficult to calculate this number practically. Hence, it is important to modify Algorithm 1 in such a way that it does not use $\phi(\mathcal{F}, S_{\mathfrak{C}})$. This is done in Algorithm 2. The following theorem is originally stated in [15] for a locally Lipschitz function.

Theorem 4.12 *Let $\mathfrak{C}$ be the set of all Clarke stationary points of problem* (2.1). *Let the lower level problem satisfy* (*MFCQ*), (*SSOC*) *and* (*CRCQ*) *and let X be compact. Assume also that $\{\rho_k\}$ is produced in such a way that $\lim_{k\to\infty} \rho_k = 1$. Then, the sequence $\{x_k\}$ produced by the Algorithm 2 has a subsequence that converges to a point in $\mathfrak{C}$.*

Proof Let the sequence $\{t_k\}$ be produced by the algorithm. From Lemma 4.9 we have $\lim_{k\to\infty} t_k = 0$. Let $\delta > 0$ be sufficiently small. Set

$$S_{\mathfrak{C}}(\delta) = X \setminus \bigcup_{x\in\mathfrak{C}} B(x,\delta).$$

Algorithm 2 Direct search algorithm for bilevel programming: the second version

Initialization: Select an initial point $x_\circ$ and a positive number $\rho_\circ < 1$. Consider constants $c_1, c_2, t_\circ, \beta_{\max}, \beta_{\min}$ with $0 < c_1 < 1 < c_2$, $\beta_{\max} \geq \beta_{\min}$. $k := 0$. Consider a continuous single variable function $\omega(t)$ such that $\lim_{t\to 0} \frac{\omega(t)}{t} = 0$ and $\omega(t) > 0$ if $t > 0$.

Step 1: Choose D_k with vector density $\rho_k > 0$ and $\beta_{\min} \leq \|d\| \leq \beta_{\max}, \forall d \in D_k$.

Step 2: Search for $d_k \in D_k$ such that $\mathcal{F}(x_k + t_k d_k) < \mathcal{F}(x_k) - \omega(t)$. If there exists such a direction d then go to Step 3 otherwise go to Step 4.

Step 3 (Successful iterate): Do the following:
set $x_{k+1} = x_k + t_k d_k$
set $t_{k+1} = c_{2_k} t_k, \quad c_{2_k} \in [1, c_2)$ go to Step 1.

Step 4 (Unsuccessful iterate): Do the following:
Set $x_{k+1} = x_k$
Set $t_{k+1} = c_{1_k} t_k, \quad c_{1_k} \in (c_1, 1)$ go to Step 5.

Step 5 (Stopping condition): If $t_{k+1} < t_\circ$ then go to Step 6.

Step 6: Select $\rho_{k+1} > \rho_k$. Pick $D_{k+1} \supseteq D_k$ with $\rho(D_{k+1}) \geq \rho_{k+1}$ and $\beta_{\min} \leq \|d\| \leq \beta_{\max}, \forall d \in D_{k+1}$. Put $k := k + 1$ and go to Step 2.

From Theorem 4.6, for every such set $S_{\mathfrak{C}}(\delta)$ we have $\phi(\mathcal{F}, S_{\mathfrak{C}}(\delta)) > 0$. Since $\rho_k \to 1$ as $k \to \infty$ there must exist a constant $k_\circ$ such that $\rho_k > \phi(\mathcal{F}, S_{\mathfrak{C}}(\delta))$ for $k \geq k_\circ$. Let $D_{k_\circ}$ be a finite generating set with vector density $\rho_{k_\circ}$. By Lemma 4.8 there exist τ and ϵ such that for all $t \in (0, \tau]$ we have

$$\min_{v \in D_{k_\circ}} \mathcal{F}(x + tv) - \mathcal{F}(x) \leq -\epsilon t.$$

For every $k \geq k_\circ$ we have $D_{k_\circ} \subseteq D_k$. That means $\min_{v\in D_k} \mathcal{F}(x + tv) - \mathcal{F}(x) \leq -\epsilon t$ for the above τ and ϵ. The remaining proof is similar to the proof of Theorem 4.11. □

5 Conclusion

We have shown that, if the lower level problem is a convex parametric optimization problem satisfying (MFCQ), (CRCQ) and (SSOC), then the bilevel programming problem can be replaced by single level Lipschitz optimization problem. The direct search method can be applied to the later problem. We have presented two versions of the algorithm. However, the algorithm developed here in this paper is purely theoretical and conceptual because for efficient implementation we need the explicit knowledge of the solution function which is difficult to determine in practice. The explicit computation of the solution function $y(\cdot)$ is left for future investigation. Moreover, in our future work we extend our study to the application of direct search methods for bilevel programming problems with coupling constraints.

Acknowledgement The authors wish to thank one anonymous referee for valuable comments which helped us to improve the presentation of this paper.

References

1. Abramson, M.A.: Second-order behavior of pattern search. SIAM J. Optim. **16**, 515–530 (2005)
2. Abramson, M.A., Audet, C.: Convergence of mesh adaptive direct search to second-order stationary points. SIAM J. Optim. **17**, 606–619 (2006)
3. Audet, C.: Convergence results for generalized pattern search algorithms are tight. Optim. Eng. **5**, 101–122 (2004)
4. Audet, C., Dennis, J.E. Jr.: Analysis of generalized pattern searches. SIAM J. Optim. **13**, 889–903 (2003)
5. Audet, C., Dennis, J.E. Jr.: A pattern search filter method for nonlinear programming with out derivatives. SIAM J. Optim. **14**, 980–1010 (2004)
6. Audet, C., Dennis, J.E. Jr.: Mesh adaptive direct search algorithms for constrained optimization. SIAM J. Optim. **17**, 188–217 (2006)
7. Bard, J.F.: Some properties of the bilevel programming problem. J. Optim. Theory Appl. **68**, 371–378 (1991)
8. Bard, J.F.: Practical Bilevel Optimization: Algorithms and Applications. Kluwer Academic, Dordrecht (1998)
9. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
10. Dempe, S.: A necessary and a sufficient optimality condition for bilevel programming problems. Optimization **25**, 341–354 (1992)
11. Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic, Dordrecht (2002)
12. Hager, W.W.: Lipschitz continuity for constrained processes. SIAM J. Control Optim. **17**, 321–269 (1979)
13. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: New perspectives on some classical and modern methods. SIAM Rev. **45**, 385–482 (2003)
14. Mifflin, R.: Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 959–972 (1977)
15. Popović, D., Teel, A.R.: Direct search methods for nonsmooth optimization. In: 43rd IEEE Conference on Decision and Control, Vol. 3, pp. 3173–3178, Atlantis, Paradise Island, Bahamas, December 14–17, 2004
16. Ralph, D., Dempe, S.: Directional derivatives of the solution of a parametric nonlinear program. Math. Program. **70**, 159–172 (1995)
17. Scholtes, S.: Introduction to piecewise differentiable equations. Preprint 53-1994, Universität Karlsruhe, Karlsruhe (1994)
18. Torczon, V.: On the convergence of the multidirectional search algorithm. SIAM J. Optim. **1**, 123–145 (1991)
19. Torczon, V.: On the convergence of pattern search algorithms. SIAM J. Optim. **7**, 1–25 (1997)
20. Von Stackelberg, H.: Marktform und Gleichgewicht. Springer, Berlin (1934). Engl. transl.: The Theory of the Market Economy. Oxford University Press (1952)