Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 03–08

A MULTICRITERIA APPROACH TO BILEVEL OPTIMIZATION

JÖRG FLIEGE AND LUÍS N. VICENTE

Abstract: In this paper we study the relationship between bilevel optimization and multicriteria optimization. Given a bilevel optimization problem, we introduce an order relation such that the optimal solutions of the bilevel problem are the nondominated points with respect to the order relation. In the case where the lower level problem of the bilevel optimization problem is convex and continuously differentiable in the lower level variables, this order relation is equivalent to a second, more tractable order relation.

Then, we show how to construct a (nonconvex) cone for which we can prove that the nondominated points with respect to the order relation induced by the cone are also nondominated points with respect to any of the two order relations mentioned before. We also comment on the practical and computational implications of our approach.

Keywords: Bilevel Optimization, Multicriteria Optimization.
AMS Subject Classification (2000): 49K35, 49K40, 90C26, 90C31, 90C33, 90C47.

1. Introduction

In a bilevel optimization problem some of the variables (the lower level variables) are constrained to be the solution of a inner or lower level problem. The remaining variables are called the upper level variables and parameterize the lower level problem. Similarly, the problem's objective function is called the upper level function. The objective function of the lower level problem is referred to as the lower level function.

Although several authors have attempted to establish a link between bicriteria optimization and bilevel optimization in the linear case (Bard [3] and Ünlü [19]), none have succeeded thus far in proposing conditions that guarantee that the optimal solution of a given bilevel program is nondominated for both upper and lower level objective functions (counter-examples were reported by Candler [6], Clarke and Westerberg [7], Haurie, Savard, and White [14], and Marcotte [15]). Wen and Hsu [21] suggested a sufficient

Support for this author was provided by Centro de Matemática da Universidade de Coimbra, by FCT under grant POCTI/35059/MAT/2000, by the European Union under grant IST-2000-26063, and by Fundação Calouste Gulbenkian. The author would also like to thank the IBM T.J. Watson Research Center and the Institute for Mathematics and Its Applications for their local support.

condition for that to happen ($d_1^\top d_2 \leq 0$, where d_1 and d_2 are the cost vectors of the upper level variables for upper and lower level objective functions, respectively) but a counter-example by Marcotte and Savard [16] showed that that was also false.

An attempt to address this issue was reported by Fülöp [12] and described in a section about the relationships between bilevel and multicriteria optimization included in the recent book by Dempe [8]. It is shown that the feasible set (sometimes called induced region) of a linear bilevel program can be written as the set of nondominated points of a multicriteria problem with the standard ordering cone with k criteria, where k is the rank of a given matrix plus 2. In our work, not only we consider nonlinear bilevel optimization problems, but we also characterize solutions of the bilevel problem (not just feasible points) as solutions of a multicriteria problem. Fülöp [12] was, however, the first who observe that more than just two criteria are needed to establish the link between bilevel optimization and multicriteria optimization.

It should not be seen as a surprise that the optimal solution of a bilevel problem can be dominated in terms of upper and lower level functions. The hierarchical nature of the two levels is, in the authors' view, a natural justification for this occurrence. Our approach differs from the ones mentioned above in the sense that we build the multicriteria optimization problem not directly from the upper and lower level objective functions but by using information from the whole bilevel optimization problem. In particular, we use the optimality of the lower level problem with respect to the lower level variables.

This paper is divided as follows. We provide the necessary background about multicriteria and bilevel optimization in Sections 2 and 3, respectively. Our multicriteria (more precisely, four-criteria) optimization approach to bilevel optimization is explained in Section 4 for *unconstrained* bilevel optimization. Section 5 covers the practical implications of this approach and considers extensions to *constrained* bilevel optimization and three-level optimization. In the last section of this paper we summarize our contribution and discuss its potential in applications of bilevel optimization with expensive function evaluations.

We use $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$. The norm $\|\cdot\|$ used in this paper is arbitrary and $\|\cdot\|_2$ denotes the ℓ_2 norm.

2. Multicriteria optimization

In this section we provide a brief introduction of the concepts in multicriteria optimization that are used later in this paper. Readers familiar with multicriteria optimization may wish to proceed directly to the next section.

In multicriteria optimization, several functions, say, $f_1, \ldots, f_m : \mathbb{R}^n \longrightarrow \mathbb{R}$ ($m > 1$), have to be minimized simultaneously. Such a problem can be conveniently stated in the form

$$\text{“min”} f : \mathbb{R}^n \longrightarrow \mathbb{R}^m, \tag{1}$$

where $f = (f_1, \ldots f_m)^\top$, and the meaning of "min" still has to be specified. The reason for this formulation is that there is no standard total order for the image space $\mathbb{R}^m$. In contrast to this situation, in the classical single-criterion case one always uses the standard total order defined by $x < y :\Longleftrightarrow y - x \in \mathbb{R}_{++}$ $(x, y \in \mathbb{R})$. Nevertheless, the idea of specifying an order by using a specific set defining it can be conveniently employed in multicriteria optimization, as the following discussion will show.

If an arbitrary order relation $\prec$ on $\mathbb{R}^m$ and a set $M \subseteq \mathbb{R}^m$ are given, the vector $a \in \mathbb{R}^m$ is called *minimal* or a *minimizer* w.r.t. $\prec$ in M if $a \in M$ and $a \preceq b$ for all $b \in M$. (Here, $\preceq$ is the reflexive hull of $\prec$, i.e., $a \preceq b$ if and only if $a = b$ or $a \prec b$.) Minimal points usually do not exist, one reason being that it is seldom the case that $\preceq$ is a total order. A weaker concept, the concept of *domination* is therefore needed. A point a *dominates* a point b, if $a \prec b$ and $a \neq b$ holds. A point a is *nondominated* in M, if $a \in M$ and there does not exist a point $c \in M$ with $c \prec a$ and $c \neq a$. This approach raises the question about which of the many orders in $\mathbb{R}^m$ one should choose when solving multicriteria problems.

Let $K \subseteq \mathbb{R}^m$ be an arbitrary set. Define the order

$$x <_K y \quad :\Longleftrightarrow \quad y - x \in K. \tag{2}$$

The next theorem is well known, e.g. [13].

Theorem 2.1. *Let $K \subset \mathbb{R}^m$ be a set and let $<_K$ be the binary relation defined by K as in (2). Then, the following statements hold:*

(1) *If $0 \in K$ then $<_K$ is reflexive.*
(2) *If $K + K \subseteq K$ then $<_K$ is transitive.*
(3) *If K is a cone containing no lines, i.e., $K \cap -K \subseteq \{0\}$ (such a cone is also called* pointed*), then $<_K$ is anti-symmetric.*
(4) *The order $<_K$ is total if and only if $K \cup -K = \mathbb{R}^m$.*

(5) *The set K is closed if and only if the relation $<_K$ is "continuous at 0" in the following sense. For all $a \in \mathrm{I\!R}^m$ and all sequences $(a^{(i)})_{i \in I\!N}$ in $\mathrm{I\!R}^m$ with $\lim_{i \to +\infty} a^{(i)} = a$ and $0 <_K a^{(i)}$ for all $i \in I\!N$ it follows that $0 <_K a$ holds.*

Note that $K + K \subseteq K$ holds if K is a convex cone. Moreover, suppose that K is a pointed cone such that the order $<_K$ is total. Then, consider $S := S^{m-1} \subset \mathrm{I\!R}^m$, the $\|\cdot\|_2$-sphere. Since $<_K$ is total, $(K \cap S) \cup (-K \cap S) = S$. But K is pointed, so for $x \in K$ we have $-x \notin K$, and therefore $(K \cap S) \cap (-K \cap S) = \emptyset$. With S endowed by the induced topology, S is an open set and can therefore not be written as the union of two disjoint closed sets. Therefore, K is not closed, a rather problematic situation with respect to numerical algorithms. According to the theorem above, practicioners prefer to choose a closed convex cone K with $0 \in K$ which contains no lines to define the partial order $<_K$. (Note that the lexicographic order in $\mathrm{I\!R}^m$ is defined by a cone which is not closed.) Moreover, in our context the space $\mathrm{I\!R}^m$ will be the image space of functions to be minimized. As a consequence, it is important for numerical reasons to have scale-invariance of the induced order. This means that if $x <_K y$ and $\lambda > 0$ then $\lambda x <_K \lambda y$, a property which holds if and only if the set K is a cone.

Using a fixed set K to define an order relation as in (2) has one additional advantage. For an arbitrary relation $\prec$, the sets

$$C(a) := \{b \in \mathrm{I\!R}^m \mid a \prec b\} - a \tag{3}$$

are constant if there exists a set K such that $\prec \, = \, <_K$ holds. Indeed, if $\prec \, = \, <_K$ then $C(a) = \{b \in \mathrm{I\!R}^m \mid a \prec b\} = \{b \in \mathrm{I\!R}^m \mid a <_K b\} = \{b \in \mathrm{I\!R}^m \mid b - a \in K\} = a + K$. This means that $\prec$ is translation-invariant, i.e., $x + z \prec y + z$ for all z if and only if $x \prec y$.

To summarize the discussion above: we are in search for a convex cone K with $0 \in K$ in order to define an order $<_K$. Other attributes of K that can be used to our advantage are closedness, pointedness, and $K \cup -K = \mathrm{I\!R}^m$ but, as pointed out before, we can not have all of these at the same time.

Usually, K is just the positive orthant without the origin, $K = \mathrm{I\!R}^m_+ \setminus \{0\}$, which gives exactly the standard definition of order in multicriteria optimization. The set of solutions (i.e., the set of efficient or Pareto points) of the problem (1) is the preimage of all nondominated points of the set M with respect to the order $<_K$.

The standard strategy to compute nondominated elements w.r.t. $<_K$ is now as follows. First, we need a technical definition, generalizing the concept of monotonicity.

Definition 2.2 (Monotonicity). *Let $K \subseteq \mathbb{R}^m$ be a cone, $M \subseteq \mathbb{R}^m$ a set, and $s : M \longrightarrow \mathbb{R}$ a function. The function s is called* K-monotonically increasing in M *if $a <_K b$ implies $s(a) \leq s(b)$ for all $a, b \in M$. The function s is called* strictly K-monotonically increasing in M *if $a <_K b, a \neq b$ implies $s(a) < s(b)$ for all $a, b \in M$.*

The set M that we will use in what follows will be the image set of the function f, i.e., $M := f(\mathbb{R}^n)$.

Functions monotone with respect to an arbitrary binary relation $\prec$ are also called *consistent with respect to* $\prec$, see [22, Chapter 1], or *order-preserving*, see [10, Chapter 7] for an overview. These functions play an important role in multicriteria optimization, as it is shown in the next theorems.

Theorem 2.3. *Let $K \subseteq \mathbb{R}^m$ be a cone with $0 \in K$ and $\{0\} \neq K \neq \mathbb{R}^m$, and let $M \subseteq \mathbb{R}^m$ be a set. Let $s : M \longrightarrow \mathbb{R}$ be a K-monotone increasing function, and let $a \in M$ be a minimum of s over M. If a is unique or if s is strictly K-monotone in M, then a is nondominated in M with respect to $<_K$.*

The proofs can be found in [20, Chapter 2] or in [13, Section 2.20].

The simplest, most widely used K-monotone functions are the linear forms in $\operatorname{int}(K^*)$, where K^* is the *dual cone* of K defined by

$$K^* = \{\omega \mid \forall a \in K : \langle \omega, a \rangle \geq 0\},$$

with $\langle \cdot, \cdot \rangle$ the (standard) inner product in $\mathbb{R}^m$. Other, nonlinear K-monotone functions have only recently attracted some attention, mainly for numerical reasons [11]. It turns out that in the case of convex cones and sets, only these linear forms need to be considered, at least in theory. More precisely, we need the linear forms from the *quasi-interior* of K^*, i.e., the set $K^+ := \{\omega \in \mathbb{R}^m \mid \forall\, a \in K \setminus \{0\} : \langle \omega, a \rangle > 0\}$. With this, it turns out that basically "all" efficient points can be found by minimizing functions of the form $\langle \omega, \cdot \rangle$ over M, as the following theorem shows.

Theorem 2.4. *Let $K, M \subseteq \mathbb{R}^m$ and define*

$$E(M, K) := \{a \in M \mid a \text{ nondominated in } M \text{ w.r.t } <_K\}$$

as well as

$$P(M,K) := \bigcup_{\omega \in K^+} \arg\min\{\langle \omega, a\rangle \mid a \in M\}.$$

Then, the following statements hold:

(1) *Let K be a convex cone with $0 \in K$ and $\{0\} \neq K \neq \mathbb{R}^m$. Then*

$$P(M,K) \subseteq E(M,K).$$

(2) *Let K be a closed convex cone with $0 \in K$ such that K contains no lines. Let M be closed and convex. Then*

$$E(M,K) \subseteq \mathrm{cl}(P(M,K)).$$

(3) *Let $K = \mathbb{R}^m_+$ and M polyhedral. Then*

$$E(M,K) = P(M,K).$$

Different proofs can be found in various textbooks and original articles. See, e.g. [13, Section 2.22], [17, Page 74], or [9]. The first proof of Part 2 is due to Arrow, Barankin, and Blackwell [2].

So, in the convex case, by varying the function s over "all" monotone functions we can generate "all" nondominant points in $f(\mathbb{R}^n)$ w.r.t. $<_K$ and their corresponding preimages by solving the problems

$$\begin{array}{ll} \min & s(f(x)) \\ \text{s.t.} & x \in \mathbb{R}^n. \end{array} \tag{4}$$

Some boundary points (more precisely, a subset of measure zero) are left out in the convex case, but this is a detail rarely of importance in applications.

It is also possible to use quadratic functions to generate nondominated points. Indeed, K-monotone quadratic functions have recently attracted some attention, mainly due to their favorable numerical properties. We start by giving a characterization of K-monotonically increasing quadratic functions. We will make use of the notation $QM := \{Qx \mid x \in M\}$, for arbitrary matrices $Q \in \mathbb{R}^{m\times m}$ and arbitrary sets $M \subseteq \mathbb{R}^m$.

Theorem 2.5. *Let $K \subseteq \mathbb{R}^m$ be a closed convex cone, let $M \subseteq \mathbb{R}^m$ be a nonempty convex set and L be the subspace parallel to* aff(M)*, the affine hull of M. Let $Q \in \mathbb{R}^{m\times m}$ be a symmetric positive semidefinite matrix and define the function v by $v(x) := \langle Qx, x\rangle$ for all $x \in M$. Then v is K-monotonically increasing on M if and only if*

$$QM \subseteq K^* + L^\perp$$

holds. Furthermore, if K is closed and $\mathrm{int}(K^* + L^\perp) \neq \emptyset$*, then v is strictly K-monotonically increasing on* $\mathrm{relint}(M)$ *if and only if*

$$Q\mathrm{relint}(M) \subseteq \mathrm{relint}(K^*) + L^\perp$$

holds.

The proof of this theorem as well as more details about K-monotonically increasing quadratic functions can be found in [11]. The next result describes one way to generate nondominated elements w.r.t. $<_K$ when considering quadratic functions [11].

Theorem 2.6. *Let K be a convex cone and M be a convex set with nonempty interior. Let the sets $E(M,K)$ and $P(M,K)$ be defined as in Theorem 2.4. For an arbitrary set $\mathcal{B}$ of matrices in $\mathbb{R}^{m\times m}$ define*

$$A(M,\mathcal{B}) := \bigcup_{Q\in\mathcal{B}} \arg\min\{\langle Qy, y\rangle \mid y \in M\}.$$

(1) *Let*

$$Q_{M,K} := \{Q \in \mathbb{R}^{m\times m} \textit{ positive definite} \mid QM \subseteq K^*\}.$$

Then

$$A(M, Q_{M,K}) \subseteq E(M,K).$$

(2) *Let $K^+ \subseteq \mathbb{R}^m_{++}$ and $M \subset \mathbb{R}^m_{++}$. Define*

$$\hat{Q}_{M,K} := \{\mathrm{diag}(\omega_1/z_1, \ldots, \omega_n/z_n) \mid z \in M, \omega \in K^+\}.$$

Then

$$P(M,K) \subseteq A(M, \hat{Q}_{M,K}).$$

3. Bilevel optimization

A common formulation for the bilevel optimization problem is as follows:

$$\begin{aligned} \min_{x_u\in\mathbb{R}^{n_u}, x_\ell\in\mathbb{R}^{n_\ell}} \quad & f_u(x_u, x_\ell) \\ \text{s.t.} \quad & g_u(x_u) \leq 0 \\ & x_\ell \in \arg\min_{g_\ell(x_u,\cdot)\leq 0} f_\ell(x_u,\cdot), \end{aligned} \tag{5}$$

where $f_u, f_\ell : \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell} \longrightarrow \mathbb{R}$ are the upper level and lower level objective functions, respectively, and $g_u : \mathbb{R}^{n_u} \longrightarrow \mathbb{R}^p$, $g_\ell : \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell} \longrightarrow \mathbb{R}^q$ define the upper level and lower level constraints, respectively.

Let us define the set of lower level minimizers for a given x_u by

$$x_\ell(x_u) = \arg\min\{f_\ell(x_u,\cdot) | \, g_\ell(x_u,\cdot) \leq 0\}.$$

The feasible region of the bilevel problem, also called the induced region, is thus defined by the set

$$\{(x_u, x_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell} | g_u(x_u) \leq 0,\ x_\ell \in x_\ell(x_u)\}\,.$$

The induced region is typically a nonconvex set even if all the functions defining it are linear. In the presence of upper level constraints of the form $g_u(x_u) \leq 0$ the induced region is a connected set. However, if we consider upper level constraints involving the lower level variables, of the form $g_u(x_u, x_\ell) \leq 0$, then the induced region could become a disconnected set.

It is possible to derive optimality conditions for the bilevel problem in the case where $x_\ell(x_u)$ is a singleton and in the case where it is not. Also, it is possible to guarantee, under appropriate conditions, that the set $x_\ell(x_u)$ is a singleton. We refer the reader to the books on bilevel optimization by Bard [4], Dempe [8], and Shimizu, Ishizuka, and Bard [18]. Moreover, if the lower level problem is convex and continuously differentiable in the lower level variables then it admits a necessary and sufficient representation in terms of its first-order necessary conditions (under the presence of a constraint qualification). The resulting problem is a (one-level) optimization problem.

4. A multicriteria approach to bilevel optimization

Now let us consider a bilevel problem with no upper or lower level constraints:

$$\begin{array}{cl} \displaystyle\min_{x_u \in \mathbb{R}^{n_u}, x_\ell \in \mathbb{R}^{n_\ell}} & f_u(x_u, x_\ell) \\ \text{s.t.} & x_\ell \in \arg\min f_\ell(x_u, \cdot) \end{array} \tag{6}$$

specified by two functions $f_u, f_\ell : \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell} \longrightarrow \mathbb{R}$. Our goal is to define a (nonreflexive) order $\prec$ that captures exactly the optimality properties of the bilevel problem, and then proceed by looking at what set might induce this order. We want to define an order $\prec$ in such a way that all solutions to the bilevel problem (6) are nondominated elements of $\mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$ with respect to $\prec$. And, of course, it would be desirable if all nondominated elements of $\prec$ are solutions to (6).

For $x = (x_u, x_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$ we have the following chain of equivalences:

$$\begin{array}{ll} x \text{ solves (6)} & \\ \iff & x \text{ feasible and} \\ & \nexists y \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell} : y \text{ feasible and } f_u(y) < f_u(x) \\ \iff & \nexists \bar{x}_\ell \in \mathbb{R}^{n_\ell} : f_\ell(x_u, \bar{x}_\ell) < f_\ell(x) \end{array}$$

$$\begin{aligned}&\text{and}\\ &\nexists y = (y_u, y_\ell) \in \mathrm{I\!R}^{n_u} \times \mathrm{I\!R}^{n_\ell} : y_\ell \in \arg\min f_\ell(y_u, \cdot)\\ &\qquad\qquad\qquad\qquad\qquad \text{and } f_u(y) < f_u(x).\end{aligned}$$

Therefore,

$$\begin{aligned}& x \text{ does not solve (6)}\\ \iff\ & \exists \bar{x}_\ell \in \mathrm{I\!R}^{n_\ell} : f_\ell(x_u, \bar{x}_\ell) < f_\ell(x)\\ & \text{or}\\ & \exists y = (y_u, y_\ell) \in \mathrm{I\!R}^{n_u} \times \mathrm{I\!R}^{n_\ell} : y_\ell \in \arg\min f_\ell(y_u, \cdot)\\ &\qquad\qquad\qquad\qquad\qquad \text{and } f_u(y) < f_u(x).\end{aligned}$$

Now let $x = (x_u, x_\ell), y = (y_u, y_\ell) \in \mathrm{I\!R}^{n_u} \times \mathrm{I\!R}^{n_\ell}$ be given and define

$$\begin{aligned} x \prec y \ \ :\iff\ & [x_u = y_u \text{ and } f_\ell(x) < f_\ell(y)]\\ & \text{or}\\ & [x_\ell \in \arg\min f_\ell(x_u, \cdot) \text{ and } f_u(x) < f_u(y)]\,.\end{aligned}$$

We then have the following result.

Theorem 4.1. *The point $z = (z_u, z_\ell) \in \mathrm{I\!R}^{n_u} \times \mathrm{I\!R}^{n_\ell}$ solves the bilevel problem (6) if and only if z is nondominated with respect to $\prec$.*

Proof: The proof follows immediately from the definition of $\prec$. In fact, if z does not solve the bilevel problem (6) then it is either because it is not feasible and there would be an $w = (w_u, w_\ell)$ such that $w_u = z_u$ and $f_\ell(w) < f_\ell(z)$, yielding $w \prec z$, or it is because there is a $w = (w_u, w_\ell)$ such that $w_\ell \in \arg\min f_\ell(w_u, \cdot)$ and $f_u(w) < f_u(z)$, also yielding $w \prec z$. The other implication is proved similarly. ■

Unfortunately, $\prec$ is a difficult relation to work with. In fact, given $C(x)$ defined as in (3), it is not clear how to get $C(x) - x = C(y) - y \ \forall x, y$. Therefore, we suggest the use of a weaker relation, $\tilde{\prec}$, defined under suitable differentiability assumptions by

$$\begin{aligned} x \tilde{\prec} y \ \ :\iff\ & [x_u = y_u \text{ and } f_\ell(x) < f_\ell(y)]\\ & \text{or}\\ & [\|D_2 f_\ell(x_u, x_\ell)\| = 0 \text{ and } f_u(x) < f_u(y)]\,,\end{aligned}$$

where $D_2 f_\ell(x_u, x_\ell)$ denotes the gradient of f_ℓ with respect to the second argument x_ℓ and $\|\cdot\|$ is an arbitrary norm (for instance $\|\cdot\|_2$). Of course, if

$f_\ell(x_u, \cdot)$ is convex for all x_u, we have that $\prec$ and $\tilde{\preceq}$ are the same. However, we always have, even in the nonconvex case, that

$$x \prec y \qquad \Longrightarrow \qquad x \,\tilde{\preceq}\, y. \tag{7}$$

So, the set of nondominated points of $\tilde{\preceq}$ is included in the set of nondominated points of $\prec$.

The question now is how can we compute the set of nondominated points of $\tilde{\preceq}$. For this purpose, we introduce the function

$$F : x = (x_u, x_\ell) \longmapsto (x_u, f_u(x), f_\ell(x), \|D_2 f_\ell(x)\|).$$

The image space of this function is $\mathbb{R}^r = \mathbb{R}^{n_u} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. In this space, we then define the cone

$$K := \{(x, f_1, f_2, d) \in \mathbb{R}^r \mid (x = 0 \text{ and } f_2 > 0) \text{ or } (f_1 > 0 \text{ and } d \geq 0)\}$$

and the induced order $<_K$ by (2). The following result provide us a scheme to compute nondominated points w.r.t. to $\tilde{\preceq}$.

Theorem 4.2. *If $F(z) \in \mathbb{R}^r$ is nondominated with respect to $<_K$, for some $z = (z_u, z_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$, then z is nondominated with respect to $\tilde{\preceq}$.*

Proof: Let us assume that there exists no $\bar{F} \in \mathbb{R}^r$ such that $\bar{F} <_K F(z)$, i.e., such that $F(z) - \bar{F} \in K$. Consequently, there exists no $F(w)$, with $w = (w_u, w_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$, such that $F(z) - F(w) \in K$, in other words there exists no w such that

$$[w_u = z_u \text{ and } f_\ell(w) < f_\ell(z)]$$
$$\text{or}$$
$$[\|D_2 f_\ell(w_u, w_\ell)\| \leq \|D_2 f_\ell(z_u, z_\ell)\| \text{ and } f_u(w) < f_u(z)]\,.$$

This in turn implies that $\|D_2 f_\ell(z_u, z_\ell)\| = 0$ because otherwise there would be an $\bar{w}$ such that $\bar{w}_u = z_u$ and $f_\ell(\bar{w}) < f_\ell(z)$ contradicting what we have just stated above. Thus, there exists no $w = (w_u, w_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$ such that

$$[w_u = z_u \text{ and } f_\ell(w) < f_\ell(z)]$$
$$\text{or}$$
$$[\|D_2 f_\ell(w_u, w_\ell)\| = 0 \text{ and } f_u(w) < f_u(z)]\,,$$

which proves that z is nondominated with respect to $\tilde{\preceq}$. ■

The following corollaries are simple consequences of Theorems 4.1 and 4.2 and implication (7).

Corollary 4.3. *If $F(z) \in \mathbb{R}^r$ is nondominated with respect to $<_K$, for some $z = (z_u, z_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$, then z is nondominated with respect to $\prec$.*

Corollary 4.4. *If $F(z) \in \mathbb{R}^r$ is nondominated with respect to $<_K$, for some $z = (z_u, z_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$, then z is an optimal solution of the bilevel problem (6).*

5. Practical aspects and extensions

Unfortunately, the cone K is not convex (but rather the union of two convex cones) and due to $\mathrm{conv}(K) = \mathbb{R}^r$, we get $K^* = \{0\}$. We are allowed to use a cone L larger than K, i.e., $K \subset L$, but of course the identities $\mathrm{conv}(L) = \mathbb{R}^r$ and $L^* = \{0\}$ follow immediately. As a consequence, the standard scalarization approach outlined in Section 2 can not be directly used here.

5.1. Scalarizations. An alternative strategy is to define the convex cones $K_1 := \{(x, f_1, f_2, d) \in \mathbb{R}^r \mid x = 0 \text{ and } f_2 > 0\}$ and $K_2 := \{(x, f_1, f_2, d) \in \mathbb{R}^r \mid f_1 > 0 \text{ and } d \geq 0\}$. Then $K = K_1 \cup K_2$. Now let s_i be a K_i-monotone function $(i = 1, 2)$ (easy to construct, since the cones K_i are convex) and define $s(x) := \min\{s_1(x), s_2(x)\}$. This function s is K-monotone and continuous, and the corresponding optimization problem (4) is equivalent to solving two optimization problems, one with objective function $s_1 \circ F$ and one with objective function $s_2 \circ F$. Let us illustrate this cone decoupling with the following example

$$\begin{aligned} \min_{x_u \in \mathbb{R}, x_\ell \in \mathbb{R}} \quad & x_\ell \\ \text{s.t.} \quad & x_\ell \in \arg\min (x_\ell - x_u^2)^2/2. \end{aligned}$$

The function $F(x_u, x_\ell)$ is defined as

$$F : x = (x_u, x_\ell) \longmapsto (x_u, x_\ell, (x_\ell - x_u^2)^2/2, (x_\ell - x_u^2)^2)$$

and thus

$$s_1(F(x_u, x_\ell)) = c_1 x_u + c_3 (x_\ell - x_u^2)^2/2, \quad s_2(F(x_u, x_\ell)) = c_2 x_\ell + c_4 (x_\ell - x_u^2)^2.$$

Here we used $\|\cdot\|_2^2$ instead of $\|\cdot\|_2$. Also, we have

$$K_1 = \{0\} \times \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R}, \quad K_2 = \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{R} \times \mathbb{R}_+.$$

The constants c_1, c_2, c_3, c_4 satisfy

$$(c_1, 0, c_3, 0) \in K_1^* = \mathbb{R} \times \{0\} \times \mathbb{R}_+ \times \{0\}$$

and

$$(0, c_2, 0, c_4) \in K_2^* = \{0\} \times \mathbb{R}_+ \times \{0\} \times \mathbb{R}_+.$$

We also have

$$M = F(\mathbb{R}, \mathbb{R}) = \mathbb{R} \times \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}_+.$$

This example shows that this cone decoupling strategy together with linear scalarizations might not be appropriate. In fact, by choosing $c_1 \neq 0$ and $c_3 \in \mathbb{R}$ arbitrary we immediately see that the function $s_1(F(x_u, x_\ell))$ is unbounded below. In this case the arg min operator in the definition of $P(M, K_1)$ in Theorem 2.4 returns the empty set. When $c_1 = 0$, $\arg\min s_1(F(x_u, x_\ell))$ coincides with the induced region $\{(x_u, x_\ell(x_u)) | x_u \in \mathbb{R}\}$ of the bilevel problem. A similar situation happens when minimizing $s_2(F(x_u, x_\ell))$: if $c_2 > 0$ then $\arg\min s_2(F(x_u, x_\ell)) = \emptyset$, otherwise $\arg\min s_2(F(x_u, x_\ell))$ is the same as $\{(x_u, x_\ell(x_u)) | x_u \in \mathbb{R}\}$. With these parameter values, the minimization of $\min\{s_1(x), s_2(x)\}$ yields points in the induced region but it is not able to identify the optimal solution $(0, 0)$ of the bilevel problem.

However, we have mentioned in the end of Section 2 that quadratic scalarizations can be used too to generate nondominated points w.r.t. to an order relation defined by a closed convex cone. To construct a K_1-monotone quadratic function on M, Theorem 2.5, Part 1, tells us that we have to look for symmetric positive semidefinite matrices Q with $QM \subseteq K_1^*$. For the sake of simplicity let us consider diagonal matrices $Q = \operatorname{diag}(q_1, q_2, q_3, q_4)$ only. Then, $q_1 \in \mathbb{R}$ can be chosen arbitrarily, we have to choose $q_2 = 0$ necessarily, $q_3 \in \mathbb{R}_+$ is arbitrary, and $q_4 = 0$ necessarily again. However, we need $q_1 \geq 0$ to get positive semidefiniteness. Likewise, K_2-monotone quadratic functions with diagonal matrices Q have to fulfill $q_1 = q_2 = q_3 = 0$, while $q_4 \in \mathbb{R}_+$ can be chosen arbitrarily. The corresponding quadratic functions on $M \subset \mathbb{R}^4$ can be written as $s_1(x, f_1, f_2, d) := q_1 x^2 + q_3 f_2^2$ and $s_2(x, f_1, f_2, d) := q_4 d^2$, respectively (with $q_1 \in \mathbb{R}_+$, $q_3 \in \mathbb{R}_+$, and $q_4 \in \mathbb{R}_+$). In the end, we arrive at scalarizations of the form

$$s_1(F(x_u, x_\ell)) = q_1 x_u^2 + q_3 (x_\ell - x_u^2)^4/4, \quad s_2(F(x_u, x_\ell)) = q_4 (x_\ell - x_u^2)^4.$$

For $q_1, q_3 \geq 0$ the first objective function is bounded below, and the second objective function is bounded below for all feasible parameter values $q_4 \geq 0$. Moreover, for $q_1, q_3 > 0$ the unique minimizer of $\min s_1 \circ F$ is the unique solution to the original bilevel problem (despite the fact that Theorem 2.6, Part 1, is not applicable here since $q_2 = q_4 = 0$ yields a non positive definite matrix Q). So, for this instance problem, by choosing "sensible" parameter values we were able to solve the bilevel problem by solving parameterized (one-level) optimization problems.

In general we would have to choose a finite set of parameters to get a finite number of (one-level) optimization problems. A natural question is what type of scalarization to use (linear, quadratic, or other) and how to choose the set of parameters. These questions are hard to answer and seem to be highly problem dependent. The parameters could be chosen *a priori* based on some information about the problem or corrected *a posteriori* after analyzing information gained during the solution of a first set of parameterized (one-level) optimization problems.

5.2. The constrained bilevel case. We consider here briefly constrained bilevel problems of the form given in (5). Let $x = (x_u, x_\ell), y = (y_u, y_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$ be given and define

$$\begin{aligned} x \prec y \quad :\Longleftrightarrow \quad & [x_u = y_u \text{ and } f_\ell(x) < f_\ell(y)] \\ & \text{or} \\ & \left[x_\ell \in \arg \min_{g_\ell(x_u, \cdot) \le 0} f_\ell(x_u, \cdot) \text{ and } f_u(x) < f_u(y) \right]. \end{aligned}$$

as well as

$$\Omega := \{x \in \mathbb{R}^{n_u} \mid g_u(x_u) \le 0\} \times \mathbb{R}^{n_\ell}.$$

Then, $z = (z_u, z_\ell) \in \mathbb{R}^{n_u} \times \mathbb{R}^{n_\ell}$ solves the bilevel problem (5) if and only if z is nondominated with respect to $\prec$ in Ω.

Using the same strategy as above, define

$$\begin{aligned} x \tilde{\prec} y \quad :\Longleftrightarrow \quad & [x_u = y_u \text{ and } f_\ell(x) < f_\ell(y)] \\ & \text{or} \\ & \left[\|\mathrm{proj}_{x_u}(D_2 f_\ell(x_u, x_\ell))\| = 0 \text{ and } f_u(x) < f_u(y) \right], \end{aligned}$$

where $\mathrm{proj}_{x_u}(\cdot)$ is the projection operator (see e.g. [5]) onto the set of feasible points of the lower level problem, $\{z \in \mathbb{R}^{n_\ell} \mid g_\ell(x_u, z) \le 0\}$. Likewise, we consider now the function F defined by

$$F : x = (x_u, x_\ell) \longmapsto (x_u, f_u(x), f_\ell(x), \|\mathrm{proj}_{x_u}(D_2 f_\ell(x))\|).$$

and it is clear that the strategy outlined above for the unconstrained case would also work for the constrained case.

5.3. The three-level case. Let us now consider three-level optimization problems of the form

$$\begin{array}{ll} \min\limits_{x_1\in\mathbb{R}^{n_1},x_2\in\mathbb{R}^{n_2},x_3\in\mathbb{R}^{n_3}} & f_1(x_1,x_2,x_3) \\ \text{s.t.} & (x_2,x_3) \text{ solve } \begin{array}{rl} \min_{\xi_2,\xi_3} & f_2(x_1,\xi_2,\xi_3) \\ \text{s.t.} & \xi_3 \in \arg\min f_3(x_1,\xi_2,\cdot). \end{array} \end{array}$$

For $x_1 \in \mathbb{R}^{n_1}$ fixed, the feasibility for the problem above is controlled by the bilevel problem

$$\begin{array}{rl} \min_{\xi_2,\xi_3} & f_2(x_1,\xi_2,\xi_3) \\ \text{s.t.} & \xi_3 \in \arg\min f_3(x_1,\xi_2,\cdot). \end{array}$$

Using the result from Section 4, we define the order $\prec_{x_1}$ on $(\mathbb{R}^{n_2}\times\mathbb{R}^{n_3})\times(\mathbb{R}^{n_2}\times\mathbb{R}^{n_3})$ for two points (x_2,x_3) and (y_2,y_3) by

$$\begin{array}{rcl} (x_2,x_3) \prec_{x_1} (y_2,y_3) & :\Longleftrightarrow & [x_2 = y_2 \text{ and } f_3(x_1,x_2,x_3) < f_3(x_1,y_2,y_3)] \\ & & \text{or} \\ & & [x_3 \in \arg\min f_3(x_1,x_2,\cdot) \\ & & \ \text{and } f_2(x_1,x_2,x_3) < f_2(x_1,y_2,y_3)]\,. \end{array}$$

This parameterized order relation could then be used to define a order relation for the original three-level problem. We give a couple of alternatives in the next paragraph but none of which seems totally satisfactory. It is important to stress the high complexity of hierarchical optimization problems and to convey the point that the jump from two-level to three-level should not be regarded as a simple increase in dimensionality.

One possibility would be to define the order relation as

$$\begin{array}{rcl} x \prec y & :\Longleftrightarrow & [x_1 = y_1 \text{ and } (x_2,x_3) \prec_{x_1} (y_2,y_3)] \\ & & \text{or } [(x_2,x_3) \text{ is nondominated w.r.t. } \prec_{x_1} \text{ and } f_1(x) < f_1(y)]\,, \end{array}$$

reflecting well the structure of the original problem but leading to a relation that is difficult to check since it involves global information ((x_2,x_3) is nondominated w.r.t. $\prec_{x_1}$). Another possibility would be to define the order relation as

$$\begin{array}{rcl} x \prec y & :\Longleftrightarrow & [x_1 = y_1 \text{ and } (x_2,x_3) \prec_{x_1} (y_2,y_3)] \\ & & \text{or } [(x_2,x_3) \prec_{y_1} (y_2,y_3) \text{ and } f_1(x_1,x_2,x_3) < f_1(y_1,y_2,y_3)]\,. \end{array}$$

Now the problem is that we are checking optimality of (y_1,y_2,y_3) against points (x_1,x_2,x_3) that are not necessarily feasible.

6. Concluding remarks and future work

Bilevel optimization problems appear in a wide range of applications [4, 8, 18] and, in particular, in engineering applications related with optimal design [1]. Many of these latter problems are defined by black-box simulation codes related with different engineering disciplines, where derivatives frequently are unavailable. The use of derivative-free methods in bilevel optimization was part of the motivation for the theoretical investigations reported in this paper. Given that there exist already several sophisticated algorithms and implementations for derivative-free (one-level) optimization, a natural approach to derivative-free bilevel optimization would be to reformulate the bilevel problem into a one-level optimization problem, to allow the application of such derivative-free techniques.

Such approach is also supported from the fact that one encounters frequently (one-level) optimization problems where derivatives are unavailable and that resulted from the linear combination of functions appearing in an multicriteria optimization problem. Our approach has the same flavor. However, since there is a hierarchical structure involved, it would be wrong to address the derivative-free bilevel optimization problem as a derivative-free bicriteria optimization problem. We have therefore investigated how to repose bilevel optimization problems as appropriate multicriteria (more precisely, four-criteria) optimization problems and how to choose the corresponding appropriate scalarizations.

We would certainly like to improve our current knowledge in many directions. As discussed in Section 5.1, the choice of type of scalarization and of scalarization parameters is quite important in practice. We hope to learn more about this opening issue by looking at particular classes of bilevel optimization problems.

References

[1] N. Alexandrov, *Multilevel methods for optimal design*, in Encyclopedia of Optimization, C. A. Floudas and P. M. Pardalos, eds., Kluwer Academic Publishers, Dordrecht, 2001, pp. 528–537.

[2] K. J. Arrow, E. W. Barankin, and D. Blackwell, *Admissible points of convex sets*, in Contributions to the Theory of Games, H. W. Kuhn and A. W. Tucker, eds., Princeton University Press, Princeton, New Jersey, 1953, pp. 481–492.

[3] J. Bard, *Optimality conditions for the bilevel programming problem*, Naval Research Logistics Quarterly, 31 (1984), pp. 13–26.

[4] J. F. Bard, *Practical Bilevel Optimization: Algorithms and Applications*, Kluwer Academic Publishers, Dordrecht, 1998.

[5] P. H. Calamai and J. J. Moré, *Projected gradient methods for linearly constrained problems*, Mathematical Programming, 39 (1987), pp. 93–116.

[6] W. Candler, *A linear bilevel programming algorithm: A comment*, Computers and Operations Research, 15 (1988), pp. 297–298.

[7] P. Clarke and A. Westerberg, *A note on the optimality conditions for the bilevel programming problem*, Naval Research Logistics, 35 (1988), pp. 413–418.

[8] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.

[9] R. Durier, *Weighting factor results in vector optimization*, Journal of Optimization Theory and Applications, 58 (1988), pp. 411–430.

[10] P. C. Fishburn, *Mathematics of Decision Theory*, Methods and Models in the Social Siences, Mouton, The Hague, Paris, 1972.

[11] J. Fliege, *Gap-free computation of Pareto-points by quadratic scalarizations*, Preprints des Instituts für Angewandte Mathematik 291, Universität Erlangen-Nürnberg, Institut für Angewandte Mathematik, Universität Erlangen-Nürnberg, 91058 Erlangen, Germany, June 17, 2002. To be published in *Mathematical Methods of Operations Research.*

[12] J. Fülöp, *On the equivalence between a linear bilevel programming problem and linear optimization over the efficient set*, Tech. Rep. WP 93-1, Laboratory of Operations Research and Decision Systems, Computer and Automation Institute, Hungarian Academy of Sciences, 1993.

[13] A. Göpfert and R. Nehse, *Vektoroptimierung — Theorie, Verfahren und Anwendungen*, vol. 74 of Mathematisch-Naturwissenschaftliche Bibliothek, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1990.

[14] A. Haurie, G. Savard, and D. White, *A note on: an efficient point algorithm for a linear two-stage optimization problem*, Operations Research, 38 (1990), pp. 553–555.

[15] P. Marcotte, *A note on bilevel programming algorithm by LeBlanc and Boyce*, Transportation Research, 22 B (1988), pp. 233–237.

[16] P. Marcotte and G. Savard, *A note on the pareto optimality of solutions to the linear bilevel programming problem*, Computers and Operations Research, 18 (1991), pp. 355–359.

[17] Y. Sawaragi, H. Nakayama, and T. Tanino, *Theory of Multiobjective Optimization*, vol. 176 of Mathematics in Science and Engineering, Academic Press, 1985.

[18] K. Shimizu, Y. Ishizuka, and J. F. Bard, *Nondifferentiable and Two–Level Mathematical Programming*, Kluwer Academic Publishers, Boston, 1997.

[19] G. Ünlü, *A linear bilevel programming algorithm based on bicriteria programming*, Computers and Operations Research, 14 (1987), pp. 173–179.

[20] W. Vogel, *Vektoroptimierung in Produkträumen*, vol. 35 of Systems in Economicss, Verlag Anton Hain, Meisenheim am Glan, 1977.

[21] U. Wen and S. Hsu, *A note on a linear bilevel programming algorithm based on bicriteria programming*, Computers and Operations Research, 16 (1989), pp. 79–83.

[22] D. J. White, *Optimality and Efficiency*, John Wiley & Sons, Chichester, 1982.

Jörg Fliege
Fachbereich Mathematik, Universität Dortmund, 44221 Dortmund, Germany, fliege@math.uni-dortmund.de.

Luís N. Vicente
Departamento de Matemática, Universidade de Coimbra, 3001-454 Coimbra, Portugal (lnv@mat.uc.pt).