Ann. Acad. Rom. Sci.
Ser. Math. Appl.
ISSN 2066-6594 Vol. 12, No. 1-2/2020

WHEN SEMIVECTORIAL BILEVEL OPTIMIZATION REDUCES TO ORDINARY BILEVEL OPTIMIZATION*

Henri Bonnel†‡

DOI https://doi.org/10.56082/annalsarscimath.2020.1-2.344

Dedicated to Dr. Vasile Drăgan on the occasion of his 70th anniversary

Abstract

The paper deals with semivectorial bilevel optimization problems. The upper level is a scalar optimization problem to be solved by the leader, and the lower level is a multiobjective optimization problem to be solved by several followers acting in a cooperative way inside the greatest coalition, so choosing among Pareto solutions. In the so-called "optimistic problem", the followers choose among their best responses (i.e. Pareto solutions) one which is the most favorable for the leader. The opposite is the "pessimistic problem", when there is no cooperation between the leader and the followers, and the followers choice among their best responses may be the worst for the leader. The paper presents a general method which allows, under certain mild hypotheses, to transform a semivectorial bilevel problem into an ordinary bilevel optimization. Some applications are given.
MSC: 49J20, 49J27, 90C29, 90C48

keywords: Multiobjective optimization, bilevel optimization, semivectorial bilevel optimization problem

*Accepted for publication in revised form on June 20, 2020

†henribonnel@gmail.com Curtin University, School of Electrical Engineering, Computing and Mathematical Sciences, Mathematics and Statistics, Kent St, Bentley WA 6102, Australia.

‡The author is grateful to the anonymous referee for his comments and for his careful reading of the manuscript.

1 Introduction

A bilevel optimization problem consists in two scalar optimization problems where one problem (called lower-level problem, or the follower problem) is embedded in the other (called upper-level problem, or the leader problem). There are two kinds of variables referred as upper-level variables and lower level-variables. The origin of bilevel optimization problems can be found in the area of game theory in the work of H.F. von Stackeleberg from 1934 [49]. A bilevel optimization problem is a hierarchical optimization, and it is not a bi-criteria optimization, where both objective functions are considered jointly and a solution is some sort of best compromise between the objectives.

Multiobjective (or multicriteria) optimization origin goes back in 19^{th} century with the economic works of Edgeworth [32] and Pareto [46], but mathematical approaches began in 1951 with the famous paper of Kuhn and Tucker [40]. There are many applications of multiobjective optimization in real life problems which led to intensive researches during the last 60 years. Dealing with several conflicting objectives, a Pareto solution (called also efficient) is such that none of the objective values can be improved further without deteriorating another.

A semivectorial bilevel optimization problem is a bilevel optimization problem where the upper level is a scalar optimization problem, and the lower level is a multiobjective optimization problem.

The lower level of the semivectorial bilevel optimization problem is a parametric multiobjective (vector) optimization problem, and may be considered as a single follower having to optimize several objectives, or as several followers each of them having to optimize one (scalar) objective. The last situation corresponds to the so called greatest coalition multiplayers game. The parameter in the lower level problem is the (vector) variable chosen by the leader, and, for each choice of the leader's variable, the followers choose a Pareto solution.

In a semivectorial bilevel optimization problem the upper level is a scalar optimization problem to be solved by the leader. The leader objective depends on two (vector) variables, one chosen by the leader, and the second one represents the response of the followers.

If for each choice of the leader the followers choose among their best responses (Pareto solutions) one which is the best for the leader, so when the followers cooperate with the leader, we deal with the so-called optimistic problem.

In the case when there is no cooperation between the leader and the followers, the leader may consider the worst scenario, i.e. the situation when, for each choice of the leader, the followers choose among their best responses one which is the most unfavorables for the leader, leading to the so-called pessimistic problem.

The study of semivectorial bilevel optimization problems in Euclidean or Hilbert spaces was initiated in [16, 10], and continued by several authors [2, 21, 31, 33, 50, 48, 28]. The case of semivectorial bilevel optimal control problems was considered in [17, 18], and a study on Riemannian manifolds has been done in [19].

The semivectorial bilevel optimization problem includes as particular cases the following problems which have been intensively studied in the last decades so we will give essentially a few earlier references,

- Optimizing a scalar function over the Pareto set (introduced in [47] and investigated in [3, 4, 6, 7, 8, 9, 1, 11, 12, 13, 14, 15, 25, 26, 27, 34, 37, 38] and [51] for a survey);
- Bilevel optimization problems where the upper level and the lower level are scalar optimization problems (e.g. [45, 43, 42, 29, 24] and [30] for an extensive bibliography).

In this paper we present a general method based on scalarization of the lower level problem to reduce the semivectorial bilevel optimization problem to an ordinary bilevel optimization problem. Scalarization applies when the lower level is convex using the weighted sum approach, or when the lower level admits a utopia point using some weighted Chebyshev norm.

2 Basics results in vector optimization

2.1 Multiobjective optimization problems

Throughout the paper X and Y are real Banach spaces and $C \subset \mathbb{R}^r$ is a convex pointed cone (i.e. $\mathbb{R}_+ C \subset C$, $C + C \subset C$, $C \cap (-C) = \{0\}$). Moreover we assume that C is a closed set in $\mathbb{R}^r$, and int $(C) \neq \emptyset$, where int (A) stands for the the topological interior of any subset $A \subset \mathbb{R}^r$.

For any $z, z' \in \mathbb{R}^r$ we consider three *preference* relations (the outcome z is preferred to the outcome z')

$$z \preccurlyeq z' \Leftrightarrow z' - z \in C; \;\; z \prec z' \Leftrightarrow iz' - z \in \text{int}\,(C); \;\; z \precneqq z' \Leftrightarrow z' - z \in C \setminus \{0\}.$$

It is obvious that

$$z \prec z' \Longrightarrow z \precneqq z' \Longrightarrow z \preccurlyeq z'.$$

Notice that $\preccurlyeq$ is a partial order relation on $\mathbb{R}^r$, i.e. a reflexive, antisymmetric and transitive binary relation. Also $\prec$ and $\precneqq$ are transitive relations.

Consider a vector function $G = (G_1, \dots, G_r) : Y \to \mathbb{R}^r$, a subset S of Y (called the feasible set), and the multiobjective optimization problem

$$\text{(MOP)} \qquad\qquad \text{MIN}_C G(y) \quad \text{s.t. } y \in S.$$

For (MOP) the point $a \in S$ is called:

- *Pareto solution* if there is no $y \in S$ such that $G(y) \precneqq G(a)$, ;
- *weakly Pareto solution* if there is no $y \in S$ such that $G(y) \prec G(a)$;
- *properly Pareto solution* if a is a Pareto solution, and if there exists a pointed convex cone K such that $C \setminus \{0\} \subset \text{int } (K)$ and a is a Pareto solution for the problem $\text{MIN}_K G(y)$ s.t. $y \in S$, in other words $G(S) \cap (G(a) - K) = \{G(a)\}$.

In the particular case $C = \mathbb{R}^r_+ := \{\lambda = (\lambda_1, \dots, \lambda_r) \in \mathbb{R}^r |\ \lambda_i \geq 0,\ i = 1, \dots, r\}$ (the Pareto cone), the previous definitions can be stated as follows.

- *Pareto solution* if there is no $y \in Y$ such that, for all $i \in \{1, \dots, r\}$, $G_i(y) \leq G_i(a)$, and $G(y) \neq G(a)$;
- *weakly Pareto solution* if there is no $y \in Y$ such that, for all $i \in \{1, \dots, r\}$, $G_i(y) < G_i(a)$;
- *properly Pareto solution in the sense of Geoffrion*[(i)] if a is a Pareto solution, and there is a real number $\mu > 0$ such that for each $i \in \{1, \dots, r\}$ and every $y \in Y$ with $G_i(y) < G_i(a)$ at least one $j \in \{1, \dots, r\}$ exists with $G_j(y) > G_j(a)$ and

$$\frac{G_i(a) - G_i(y)}{G_j(y) - G_j(a)} \leq \mu.$$

We denote the set of all Pareto (resp. weakly Pareto and properly Pareto) solutions by $\underset{y \in S}{\text{ARGMIN}}_C G(y)$ (resp. w-$\underset{y \in S}{\text{ARGMIN}}_C G(y)$ and p-$\underset{y \in S}{\text{ARGMIN}}_C G(y)$). In the sequel, in order to simplify the notations, we will

[(i)] If $G(Y) + \mathbb{R}^r_+$ is convex this definition is equivalent to the general one given above

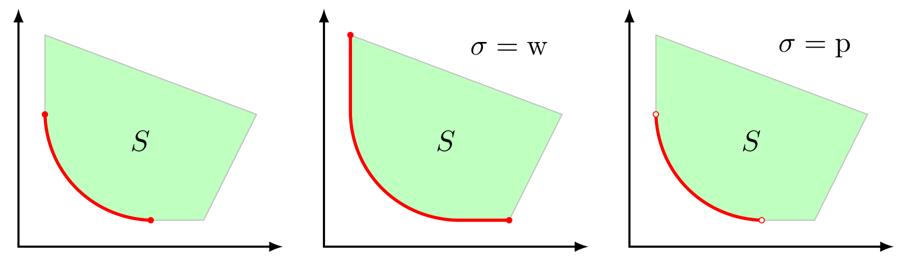

Figure 1: The figure visualizes the different sets of Pareto solutions. With some non-empty set $S \in \mathbb{R}^2$, let $G\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $G(y) = y$, and consider $C = \mathbb{R}^2_+$. In red, the figures highlight **(left)** the (image) set of all *Pareto* points; **(middle)** the set of all *weakly Pareto* points; and **(right)** the set of all *properly Pareto* points.

use the symbol $\sigma \in \{w, p\}$ for *weak* (if $\sigma = w$) or *proper* (if $\sigma = p$), i.e. we write $\sigma\text{-}\underset{y\in S}{\mathrm{ARGMIN}}_C G(y)$ for the weakly or properly Pareto solutions set.

Obviously

$$p\text{-}\underset{y\in S}{\mathrm{ARGMIN}}_C G(y) \subset \underset{y\in S}{\mathrm{ARGMIN}}_C G(y) \subset w\text{-}\underset{y\in S}{\mathrm{ARGMIN}}_C G(y) \quad (1)$$

The *vector valued function* $G = (G_1, \ldots, G_r) : Y \to \mathbb{R}^r$ is called C*-convex* if for any two points a and b in Y, we have

$$\forall t \in]0,1[\qquad G\Big((1-t)a + tb\Big) \preccurlyeq (1-t)G(a) + tG(b), \quad (2)$$

In the case $C = \mathbb{R}^r_+$ it is easy to see that

G is $\mathbb{R}^r_+$-convex if, and only if, G_i is convex for all $i = 1, \ldots, r$.

Definition 1. The problem (MOP) is called *convex* if G is C-convex and the set S is convex.

Throughout the paper $\mathbb{R}^r$ is considered with its usual euclidean structure and identified to its dual space, and we denote by $\langle \cdot, \cdot \rangle$ its usual inner product (which coincides with the duality product with our identification) and by $\| \cdot \|$ the induced norm.

The *dual cone of* C (or positive polar cone) is the set

$$C^* := \{\lambda \in \mathbb{R}^r \,|\, \langle \lambda, y\rangle \geq 0 \quad \forall y \in C\},$$

and its *quasi-interior* is given by

$$C^*_\sharp := \{\lambda \in \mathbb{R}^r \,|\, \langle \lambda, y\rangle > 0 \quad \forall y \in C \setminus \{0\}\}.$$

Notice that $(\mathbb{R}^r_+)^* = \mathbb{R}^r_+$, and $(\mathbb{R}^r_+)^*_\sharp = \text{int}\,(\mathbb{R}^r_+) = \{\lambda \in \mathbb{R}^r \,|\, \lambda_i > 0 \quad i = 1, \ldots, r\}$.

Let us denote

$$\Lambda^\sigma = \begin{cases} \{\lambda \in C^* |\ \|\lambda\|_1 = 1\} & \text{if } \sigma = w \\ \\ C^*_\sharp & \text{if } \sigma = p. \end{cases} \tag{3}$$

Proposition 1. *[19] A. The dual cone C^* is a closed set in $\mathbb{R}^r$.*

B. The set $C^*_\sharp$ (the quasi-interior of C^*) is a nonempty open set,(ii) *and it is in fact the topological interior of C^*.*

C. The set Λ^w is compact.

2.2 Scalarization

Next some scalarization theorems will be presented. Such results allow to replace a vector optimization problem with a family of scalar optimization problems. Their proofs and more details can be found in the monographs [33, 39, 41, 44, 52].

Theorem 1. (Weighted sum approach) *For each $\sigma \in \{w, p\}$ problem (MOP) satisfies*(iii)

$$\bigcup_{\lambda \in \Lambda_\sigma} \arg\min_S \langle \lambda, G\rangle \subset \sigma\text{-}\underset{y \in S}{\text{ARGMIN}}_C G(y). \tag{4}$$

Moreover, if (MOP) is convex, then the previous inclusion becomes an equality, i.e.

$$\sigma\text{-}\underset{y \in S}{\text{ARGMIN}}_C G(y) = \bigcup_{\lambda \in \Lambda_\sigma} \arg\min_S \langle \lambda, G\rangle. \tag{5}$$

(ii) This fact it is not true in general, i.e. when C is a cone in a topological vector space, but in our setting we take advantage of the finite dimension of $\mathbb{R}^r$.

(iii) $\langle \lambda, G(y)\rangle$ stands for $\sum_{i=1}^r \lambda_i G_i(y)$, and $\arg\min_S \langle \lambda, G\rangle$ denotes the set of all minimizers of the real valued function $y \mapsto \langle \lambda, G(y)\rangle$ over the set S.

Remark 1. *From (1) it is obvious that a convex (MOP) satisfies*

$$\bigcup_{\lambda\in\Lambda_p} \arg\min_S \langle\lambda, G\rangle \subset \mathrm{ARGMIN}_{y\in S\,C} G(y) \subset \bigcup_{\lambda\in\Lambda_w} \arg\min_S \langle\lambda, G\rangle \tag{6}$$

Now we will present a scalarization result for non convex (MOP).

Lemma 1. *(see [36, 39]) Let $\lambda \in \mathrm{int}\,(C)$, and denote*

$$[-\lambda, \lambda] := \{z \in \mathbb{R}^r |\, -\lambda \preccurlyeq z \preccurlyeq \lambda\},$$

in other words $[-\lambda,\lambda] = (\lambda - C) \cap (-\lambda + C)$.

Then the Minkowski functional

$$z \mapsto \|z\|_\lambda := \inf\{\alpha > 0 |\, \frac{1}{\alpha} z \in [-\lambda, \lambda]\,\} \tag{7}$$

is a norm on $\mathbb{R}^r$.

Moreover its closed unit ball coincides with $[-\lambda,\lambda]$, i.e.

$$\{z \in \mathbb{R}^r |\, \|z\|_\lambda \leq 1\} = [-\lambda, \lambda].$$

In the particular case $C = \mathbb{R}^r_+$, we have

$$\|z\|_\lambda = \max_{1\leq i\leq r} \frac{|z_i|}{\lambda_i} \tag{8}$$

which becomes the Chebyshev norm for $\lambda = (1, \ldots, 1)$.

Definition 2. We say that G *is bounded from below over* S if there exists $\hat{z} \in \mathbb{R}^r$ such that, for each $y \in S$,

$$\hat{z} \preccurlyeq G(y).$$

In other words, $G(S) \subset \hat{z} + C$.

Since the cone C is convex with nonempty interior it is easy to see that $C + \mathrm{int}\,(C) = \mathrm{int}\,(C)$. Therefore, in the definition above, take an arbitrary $e \in \mathrm{int}\,(C)$ and consider $\tilde{z} = \hat{z} - e$. We have

$$\hat{z} + C = \tilde{z} + e + C \subset \tilde{z} + \mathrm{int}\,(C).$$

So, we can state the following.

REMARK 2. *If G is bounded from below over S we can choose $\hat{z} \in \mathbb{R}^r$ such that*

$$G(S) \subset \hat{z} + \text{int }(C).$$

The following scalarization result holds.

Theorem 2. *[39] Suppose that G is C-bounded from below over S, and let $\hat{z} \in \mathbb{R}^r$ verifying*

$$G(S) \subset \hat{z} + \text{int }(C).$$

Then

$$w\text{-}\underset{y \in S}{\text{ARGMIN}}_C G(y) = \bigcup_{\lambda \in \text{int }(C)} \arg\min_{y \in S} \|G(y) - \hat{z}\|_\lambda. \tag{9}$$

In the particular case of the Pareto cone we have therefore the following.

Corollary 1. *Let $C = \mathbb{R}^r_+$, and assume that G is bounded from below on S, i.e. there exists a point $\hat{z} = (\hat{z}_1, \ldots, \hat{z}_r) \in \mathbb{R}^r$ verifying for each $i \in \{1, \ldots, r\}$*

$$\hat{z}_i < \inf\{G_i(y) | \, y \in S\}.$$

Then

$$w\text{-}\underset{y \in S}{\text{ARGMIN}}_{\mathbb{R}^r_+} G(y) = \bigcup_{\lambda \in \text{int }(\mathbb{R}^r_+)} \arg\min_{y \in S} \max_{1 \le i \le r} \lambda_i^{-1}(G_i(y) - \hat{z}_i) \tag{10}$$

Since there exists some others methods of scalarization (see [33]), which will not be considered in this paper, we prefer to give a general definition useful for the next section.

DEFINITION 3. A function $\xi : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}$ will be called *a complete scalarizing function for problem* (MOP) if there exists a *scalarizing set* $\Lambda \subset \mathbb{R}^r$ such that

$$w\text{-}\underset{y \in S}{\text{ARGMIN}}_C G(y) = \bigcup_{\lambda \in \Lambda} \arg\min_{y \in S} \xi(\lambda, G(y)), \tag{11}$$

or

$$p\text{-}\underset{y \in S}{\text{ARGMIN}}_C G(y) = \bigcup_{\lambda \in \Lambda} \arg\min_{y \in S} \xi(\lambda, G(y)). \tag{12}$$

3 The semivectorial bilevel problem

3.1 Statement of the problem

Let us call the Banach space X *the leader decision variables space* and the Banach space Y *the followers decision variables space.* Let also $f : X \times Y \to \mathbb{R}$ be the leader objective function, and let $F = (F_1, \dots, F_r) : X \times Y \to \mathbb{R}^r$ be the followers multiobjective function. Let us consider a set valued map $T : X \rightrightarrows Y$ (whose graph[(iv)] is the leader feasible set), and a set valued function $S : X \rightrightarrows Y$ (for each $x \in \text{dom }(S)$[(v)], $S(x)$ stands for the feasible follower set). We suppose that

$$\emptyset \neq \text{dom }(T) \subset \text{dom }(S).$$

Notice that the projection onto X of $\text{Gr}\,(S)$ is $\text{dom }(S)$, i.e.

$$Pr_X(\text{Gr}\,(S)) = \text{dom }(S).$$

Also

$$Pr_Y(\text{Gr}\,(S)) = \text{Range}(S) := \bigcup_{x \in X} S(x).$$

For each choice $x \in \text{dom }(T)$ the follower solves the multiobjective problem

$$(\text{MOP})_x \qquad \text{MIN}_C F(x, y) \quad \text{s.t. } y \in S(x) \cap T(x).$$

For all $x \in \text{dom }(T)$, we denote by $\psi(x)$ the *weakly or properly Pareto solution set of the follower multiobjective optimization problem,* i.e.

$$\psi(x) := \sigma\text{-}\underset{y \in S(x) \cap T(x)}{\text{ARGMIN}_C} F(x, y).$$

Thus $\psi : X \rightrightarrows Y$ is a set valued function, and *we suppose that* $\text{dom }(T) \subset \text{dom }(\psi)$, i.e., problem $(\text{MOP})_x$ admits solutions for all $x \in \text{dom }(T)$. We

(iv) The *graph* of T is the set

$$\text{Gr}\,(T) := \{(x, y) \in X \times Y \,|\, y \in T(x)\} \subset X \times Y.$$

(v)

$$\text{dom }(S) := \{x \in X \,|\, S(x) \neq \emptyset\}$$

deal with two semivectorial bilevel problems[(vi)].

- The *"optimistic semivectorial bilevel problem"*

$$(OSB) \qquad \min_{x \in \text{dom }(T)} \min_{y \in \psi(x)} f(x, y).$$

  In this case the follower cooperates with the leader, i.e., for each $x \in T$, the follower chooses amongst all its σ-Pareto solutions (his best responses) one which is the best for the leader.

- The *"pessimistic semivectorial bilevel problem"*

$$(PSB) \qquad \min_{x \in \text{dom }(T)} \max_{y \in \psi(x)} f(x, y).$$

  In this case there is no cooperation between the leader and the follower, and the leader expects the worst scenario, i.e., for each $x \in T$, the follower may choose amongst all its σ-Pareto solutions (his best responses) one which is unfavourable for the leader.

Remark 3. *Let us consider the particular case T constant, Then (OSB) represent a* strong Stackelberg game, *and (PSB) a* weak Stackelberg game *(see [22]). If, for all $x \in$* dom (T)*, there is a unique choice $y(x) \in \psi(x)$ of the followers among their best solutions, then we deal with a Stackelberg game (see [49])*

$$\min_{x \in \text{dom }(T)} f(x, y(x)).$$

Remark 4. *Let us consider the particular case $X = \{0\}$, with* dom $(T) = \{0\}$*, and denote $\tilde{f}(\cdot) = f(0, \cdot)$, $\tilde{F}(\cdot) = F(0, \cdot)$, $\tilde{S} = S(0)$, $\tilde{\psi} = \psi(0)$. Then the (OSB) becomes the problem of* minimizing over the Pareto set *:*

$$\min_{y \in \tilde{\psi}} \tilde{f}(y),$$

and (PSB) becomes the problem of maximizing over the Pareto set *:*

$$\max_{y \in \tilde{\psi}} \tilde{f}(y),$$

[(vi)] To simplify presentation and point out the main ideas we suppose that everywhere we use the symbols "min" or "max" the associated problems admit solutions, in other words the associated argmin or argmax sets are non empty. Of course this happens under some continuity and compactness (or coercivity) hypotheses.

where $\tilde{\psi}$ is the σ-Pareto set associated to the multiobjective optimization problem

$$\mathrm{MIN}_{y\in\tilde{S}}\tilde{F}(y).$$

Thus, semivectorial bilevel optimization encompasses the so-called post Pareto optimization problems*.*

3.2 A useful equivalent form

Theorem 3. *Let, for each $x \in \mathrm{dom}\ (T)$, ξ_x be a complete scalarizing function for problem* $(\mathrm{MOP})_x$*, and let Λ_x be the associated scalarizing set. Then problem (OSB) is equivalent to the following scalar bilevel optimization problem*

$$\min_{x\in\mathrm{dom}\ (T)}\left(\min_{\lambda\in\Lambda_x}\left(\min_{y\in\mathrm{argmin}_{S(x)\cap T(x)}\xi_x(\lambda,F(x,\cdot))} f(x,y)\right)\right) \tag{13}$$

On the other hand, problem (PSB) is equivalent to the following scalar min-max problem

$$\min_{x\in\mathrm{dom}\ (T)}\left(\max_{\lambda\in\Lambda_x}\left(\max_{y\in\mathrm{argmin}_{S(x)\cap T(x)}\xi_x(\lambda,F(x,\cdot))} f(x,y)\right)\right) \tag{14}$$

The equivalence is understood in the sense that for each optimal solution $(x^*,y^*) \in \mathrm{Gr}\,(T)$ of (OSB) (resp. (PSB)) there exists $\lambda^* \in \Lambda_{x^*}$ such that (x^*,λ^*,y^*) is an optimal solution of (13) (resp. (14)). Conversely, if (x^*,λ^*,y^*) is an optimal solution of (13) (resp. (14)), then $(x^*,y^*) \in \mathrm{Gr}\,(T)$ is an optimal solution of (OSB) (resp. (PSB)).

Proof. Let the function $\varphi : \{(x,\lambda) \in X \times \mathbb{R}^p |\, x \in \mathrm{dom}\ (T),\ \lambda \in \Lambda_x\} \to \mathbb{R}$ be given by

$$\varphi(x,\lambda) = \min_{y\in\mathrm{argmin}_{S(x)\cap T(x)}\xi_x(\lambda,F(x,\cdot))} f(x,y). \tag{15}$$

Obviously

the problem $\displaystyle\min_{x\in\mathrm{dom}\ (T)}\min_{\lambda\in\Lambda_x}\varphi(x,\lambda)$ *and problem (13) have the same optimal value.* (16)

On the other hand, for each $x \in \mathrm{dom}\ (T)$ the function ξ_x verifies Definition 3, therefore

$$\forall x \in T \qquad \min_{\lambda \in \Lambda_x} \varphi(x,\lambda) = \min_{y \in \psi(x)} f(x,y). \tag{17}$$

Thus

problem (OSB) *and the problem* $\min_{x \in \mathrm{dom}\,(T)} \min_{\lambda \in \Lambda_x} \varphi(x,\lambda)$ *have the same optimal value.* (18)

Hence

problem (OSB) *and problem* (13) *have the same optimal value.* (19)

Now, let $(x^*, y^*) \in \mathrm{dom}\,(T) \times S(x^*)$ be an optimal solution of (OSB).

By the hypothesis about ξ_{x^*} there exists some $\lambda^* \in \Lambda_{x^*}$ such that $y^* \in \mathrm{argmin}_{S(x^*) \cap T(x^*)} \xi_{x^*}(\lambda^*, F(x^*, \cdot))$.
The definition of φ and the fact that $y^* \in \mathrm{argmin}_{S(x^*) \cap T(x^*)} \xi_{x^*}(\lambda^*, F(x^*, \cdot))$ imply $\varphi(x^*; \lambda^*) \leq f(x^*, y^*)$. By (18) we get $\varphi(x^*; \lambda^*) \geq f(x^*, y^*)$. Hence $\varphi(x^*; \lambda^*) = f(x^*, y^*)$, which implies immediately that (x^*, λ^*, y^*) is an optimal solution for problem (13).

Conversely, let $(\hat{x}, \hat{\lambda}, \hat{y})$ be an optimal solution of problem (13).

By (15) we obtain that $\varphi(\hat{x}, \hat{\lambda})$ is the optimal value of problem (13) and $\varphi(\hat{x}, \hat{\lambda}) = f(\hat{x}, \hat{y})$.

Then, by (18) we get that $(\hat{x}, \hat{y})$ is an optimal solution of problem (OSB). Hence problems (OSB) and (13) are equivalent.

Now, in order to prove the equivalence between problem (PSB) and (14), we may consider the function $\eta : \{(x, \lambda) \in X \times \mathbb{R}^p |\, x \in \mathrm{dom}\,(Tt,\ \lambda \in \Lambda_x\} \to \mathbb{R}$ given by

$$\eta(x,\lambda) = \max_{y \in \mathrm{argmin}_{S(x) \cap T(x)} \xi_x(\lambda, F(x,\cdot))} f(x,y). \tag{20}$$

Obviously

the problem $\min_{x \in \mathrm{dom}\,(T)} \max_{\lambda \in \Lambda_x} \eta(x,\lambda)$ and problem (14) have the same optimal value. (21)

Using similar arguments as for (17) we get

$$\forall x \in \mathrm{dom}\,(T) \qquad \max_{\lambda \in \Lambda_x} \eta(x,\lambda) = \max_{y \in \psi(x)} f(x,y), \tag{22}$$

hence

$$\textit{problem (PSB) and the problem } \min_{x\in\mathrm{dom}\,(T)} \max_{\lambda\in\Lambda_x} \varphi(x,\lambda) \textit{ have the same optimal value.} \tag{23}$$

Thus

$$\textit{problem (PSB) and problem (14) have the same optimal value.} \tag{24}$$

Then, using similar arguments as for equivalence between problems (OSB) and (13) we obtain easily the conclusion. □

Next obvious Proposition presents an important particular case.

Proposition 2. *Suppose that, for each $x \in \mathrm{dom}\ (T)$, $\lambda \in \Lambda_x$, the set $\mathrm{argmin}_{S(x)\cap T(x)}\xi_x(\lambda, F(x,\cdot))$ is a singleton denoted $y(x,\lambda)$.*

Then, problems (OSB) and (PSB) respectively become

$$(\widetilde{OSB}) \qquad \min_{x\in\mathrm{dom}\,(T)} \min_{\lambda\in\Lambda_x} f(x, y(x,\lambda))$$

and

$$(\widetilde{PSB}) \qquad \min_{x\in\mathrm{dom}\,(T)} \max_{\lambda\in\Lambda_x} f(x, y(x,\lambda))$$

Remark 5. *Consider the following (ordinary) bilevel optimization problem. The leader has the same objective f, feasible set T and decision variables x as defined in this section.*

Now, the follower objective is the scalar function $g : \{(x,y,\lambda)|\,(x,y) \in \mathrm{dom}\ (T),\ \lambda \in \Lambda_x\} \to \mathbb{R}$, given by

$$g(x,y,\lambda) = \xi_x(\lambda, F(x,y)).$$

The follower decision variables are (y,λ).

For each choice of $x \in \mathrm{dom}\ (T)$ by the leader, the follower can choose any $\lambda \in \Lambda_x$, and then he solves the problem

$$\min_{y\in S(x)\cap T(x)} g(x,y,\lambda).$$

The follower best response set is

$$\bigcup_{\lambda\in\Lambda_x} \mathrm{argmin}_{y\in S(x)\cap T(x)} g(x,y,\lambda).$$

Then it is obvious that $(\widetilde{OSB})$, (resp. $(\widetilde{PSB})$) is a usual bilevel optimistic problem (resp. pessimistic problem).

4 Some examples

4.1 An illustrative example

With the notations of the previous section, let $X = \mathbb{R}$, Y a real Hilbert space with the scalar product denoted $\langle \cdot, \cdot \rangle$ and the associated norm $\| \cdot \|$. Let T, $S : \mathbb{R} \rightrightarrows Y$ with dom $(T) =$ dom $(S) =]0, +\infty[$, given for each $x > 0$ by $T(x) = S(x) = Y$.

The leader objective $f : \mathbb{R} \times Y \to \mathbb{R}$ is given for all $(x, y) \in$ dom $(T) \times Y$ by

$$f(x, y) = x + \frac{x}{2} \| \, y \|^2.$$

The followers objectives $(F_1, F_2) : \mathbb{R} \times Y \to \mathbb{R}^2$ are given for all $(x, y) \in$ dom $(S) \times Y$ by

$$(F_1(x, y), F_2(x, y)) = \frac{1}{2}(\|xy - a\|^2, \|xy - b\|^2),$$

where $a \in Y \setminus \{0\}$ (resp. $b \in Y \setminus \{0\}$) are two orthogonal vectors representing the wishes of follower #1 (resp. follower #2).

We consider the ordering cone $C = \mathbb{R}^2_+$ (the Pareto cone).

Since, for each fixed $x > 0$, $F_1(x, \cdot)$ and $F_2(x, \cdot)$ are convex functions, by Theorem 1 we have that, for each fixed $x > 0$, the properly Pareto set associated to the bi-objective problem

$$\mathrm{MIN}_{\mathbb{R}^2_+} (F_1(x, \cdot),\, F_2(x, \cdot))$$

is given by

$$\begin{aligned}
&\bigcup_{\lambda_1 > 0, \lambda_2 > 0} \operatorname{argmin}_{y \in Y} \frac{1}{2}\Big(\lambda_1 \|xy - a\|^2 + \lambda_2 \|xy - b\|^2\Big) \\
&= \bigcup_{0 < \theta < 1} \operatorname{argmin}_{y \in Y} \frac{1}{2}\Big(\theta \|xy - a\|^2 + (1 - \theta) \|xy - b\|^2\Big).
\end{aligned}$$

The last equality is obviously obtained by a division with $\lambda_1 + \lambda_2$ and putting $\theta = \frac{\lambda_1}{\lambda_1 + \lambda_2}$.

On the other hand, since the function $y \mapsto \frac{1}{2}\Big(\theta \|xy - a\|^2 + (1 - \theta) \|xy - b\|^2\Big)$ is strictly convex for fixed $\theta \in]0, 1[$ and $x > 0$, its global minimizer is given by its stationary point, hence

$$\operatorname{argmin}_{y \in Y} \frac{1}{2}\Big(\theta \|xy - a\|^2 + (1 - \theta) \|xy - b\|^2\Big) = \Big\{ \frac{1}{x}(\theta a + (1 - \theta) b) \Big\}. \quad (25)$$

Denote $y(x,\theta) = \frac{1}{x}(\theta a + (1-\theta)b)$, and by Proposition 2 the leader problem in the optimistic case becomes

$$\min_{(x,\theta)\in]0,+\infty[\times]0,1[} f(x, y(x,\theta)) = \min_{(x,\theta)\in]0,+\infty[\times]0,1[} \Big(x + \frac{1}{2x}(\theta^2\|a\|^2 + (1-\theta)^2\|b\|^2)\Big).$$

Some simple computations shows that the hessian matrix of the function $(x,\theta) \mapsto x + \frac{1}{2x}(\theta^2\|a\|^2 + (1-\theta)^2\|b\|^2)$ is positive definite for each $(x,\theta) \in]0,+\infty[\times]0,1[$. Therefore this function is strictly convex, and the unique global minimizer is its stationary point

$$\Big(x = \sqrt{\frac{1}{2}(\theta^2\|a\|^2 + (1-\theta)^2\|b\|^2)}, \theta = \frac{\|b\|^2}{\|a\|^2 + \|b\|^2}\Big)$$
$$= \Big(\frac{\|a\| \cdot \|b\|}{\sqrt{2}\|a+b\|}, \frac{\|b\|^2}{\|a+b\|^2}\Big).$$

4.2 An application in machine learning

We present an application of the optimization over the Pareto set problem (see Remark 4) to the *parameter tuning for the elastic net problem.* Consider three objectives $(F_1, F_2, F_3) : \mathbb{R}^n \to \mathbb{R}^3$ given by[(vii)]

$$x \mapsto (F_1(x), F_2(x), F_3(x)) := \Big(\frac{1}{2}\|Ax - b\|_2^2,\ \frac{1}{2}\|x\|_2^2,\ \|x\|_1\Big),$$

where $A \in \mathbb{R}^{p\times n}$ and $b \in \mathbb{R}^p$ are the *training data.* So, the lower level is the three objectives minimization problem

$$\mathrm{MIN}_{\mathbb{R}^3_+}(F_1(x), F_2(x), F_3(x)) \qquad \text{s.t. } x \in \mathbb{R}^n. \tag{26}$$

Consider also the scalar valued function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$x \mapsto f(x) := \|\tilde{A}x - \tilde{b}\|,$$

where $\tilde{A} \in \mathbb{R}^{p\times n}$ and $\tilde{b} \in \mathbb{R}^p$ are the *validation data.*

So the upper level problem is to solve

$$\min f(x) \qquad \text{s.t. } x \text{ is a properly Pareto solution for problem (26).} \tag{27}$$

[(vii)] $\|\cdot\|_p$ denotes the L^p norm, $p \geq 1$.

By Theorem 1 (with $C = \mathbb{R}^3_+$) we have that x is a properly Pareto solution for problem (26) iff there exist $\lambda_1 > 0$, $\lambda_2 > 0$, $\lambda_3 > 0$ such that x is a minimizer for the scalar function $\lambda_1 F_1 + \lambda_2 F_2 + \lambda_3 F_3$. Scaling the last function by λ_1 we can assert that

x is a properly Pareto solution for problem (26) iff there exist $\alpha > 0$, $\beta > 0$ such that x minimizes the function $x \mapsto \frac{1}{2}\|Ax - b\|_2^2 + \frac{\alpha}{2}\|x\|_2^2 + \beta\|x\|_1$.

Since, for each fixed $\alpha > 0, \beta > 0$, the last function is strictly convex and coercive, it admits a unique global minimizer $x(\alpha, \beta)$ which is called *the solution of the elastic net problem.* This problem is a good regression model, the term with L^2 norm produces regularity, and the term with L^1 norm produces sparsity. A crucial problem is the choice of parameters $\alpha > 0$ and $\beta > 0$ such that $x(\alpha, \beta)$ minimizes the error in the validation function, i.e. we want to solve the problem

$$\min_{\alpha>0,\, \beta>0} f(x(\alpha, \beta)),$$

which is exactly problem (27). For more details we refer the reader to the paper [20] where an exact algorithm to compute the optimal path $\beta \mapsto x(\alpha, \beta)$ for each fixed $\alpha > 0$ is proposed. Also it is given a formula to find β^* which minimizes the function $\beta \mapsto f(\alpha, \beta)$ (for fixed α), and finally problem (27) is solved using a grid search. Moreover, some real world examples are approached numerically in the paper [20].

Dedication

Dedicated to my friend Vasile Drăgan on his 70^{th} birthday. I am very pleased to remember that, long time ago, my first published paper was written jointly with Vasile Drăgan (see [5]).

References

[1] L.T.H. An, P.D. Tao, L.D. Muu. Numerical solution for optimization over the efficient set by D.C. optimization algorithms. *Oper. Res. Lett.* 19:117-128, 1996.

[2] Z. Ankhili, A. Mansouri. An exact penalty on bilevel programs with linear vector optimization lower level. *Eur. J. Oper. Res.* 197:36-41, 2009.

[3] H.P. Benson. Optimization over the efficient set. *J. Math. Anal.Appl.* 98:562-580, 1984.

[4] H.P. Benson. A finite, non-adjacent extreme point search algorithm for optimization over the efficient set. *J. Optim. Theory Appl.* 73:47-64, 1992.

[5] S. Bolintineanu, V. Drăgan. Open-loop iterative solution of the linear quadratic optimization problem. *Rev. Roumaine Sci. Tech. Sér. Électrotech. Énergét.* 22:535-540, 1977.

[6] S. Bolintinéanu. Optimality conditions for Minimization over the (Weakly or Properly) Efficient Set. *J. Math. Anal. Appl.* 173:523-541,1993.

[7] S. Bolintinéanu. Minimization of a quasi-concave function over an efficient set. *Math. Program.* 61:89-110, 1993.

[8] S. Bolintinéanu. Necessary conditions for nonlinear suboptimization over the weakly-efficient set. *J. Optim. Theory Appl.* 78: 579-598, 1993.

[9] S. Bolintinéanu, M. El Maghri. Pénalisation dans l'optimisation sur l'ensemble faiblement efficient. *RAIRO Oper. Res.* 31:295-310, 1997.

[10] H. Bonnel. Optimality Conditions for the Semivectorial Bilevel Optimization Problem. *Pacific Journal of Optimization,* 2:447-468, 2006.

[11] H. Bonnel. Post-Pareto Analysis For Multiobjective Parabolic Control Systems. *Ann. Acad. Rom. Sci. Ser. Math. Appl.* 5:13-34, 2013.

[12] H. Bonnel, J. Collonge. Stochastic Optimization over a Pareto Set Associated with a Stochastic Multiobjective Optimization Problem. *J. Optim. Theory Appl.* 162:405-427, 2014.

[13] H. Bonnel, J. Collonge. Optimization over the Pareto Outcome set associated with a Convex Bi-Objective Optimization Problem: Theoretical Results, Deterministic Algorithm and Application to the Stochastic case. *J. Global Optim.* 62:481-505,2015.

[14] H. Bonnel, C.Y. Kaya. Optimization Over the Efficient Set in Multiobjective Convex Optimal Control Problems. *J. Optim. Theory Appl.* 147:93-112, 2010.

[15] H. Bonnel, N.S.Pham. Nonsmooth optimization over the (weakly or properly) Pareto set of a linear- quadratic multi-objective control problem: explicit optimality conditions. *J. Ind. Manage. Optim.* 7:789-809, 2011.

[16] H. Bonnel, J. Morgan. Semivectorial Bilevel Optimization Problem: Penalty Approach. *J. Optim. Theory Appl.*131:365-382, 2006.

[17] H. Bonnel, J. Morgan. Semivectorial Bilevel Convex Optimal Control Problems: An Existence Result. *SIAM J. on Control and Optim.* 50:3224-3241, 2012.

[18] H. Bonnel, J. Morgan. Optimality Conditions for Semivectorial Bilevel Convex Optimal Control Problems. In *Computational and Analytical Mathematics In Honor of Jonathan Borwein's 60th Birthday.* Springer, 2013.

[19] H. Bonnel, L. Todjihounde, C. Udrişte. Semivectorial Bilevel Optimization on Riemannian Manifolds. *J. Optim. Theory Appl.* 167:464-486, 2015.

[20] H. Bonnel, C. Schneider. Post-Pareto analysis and a new algorithm for the optimal parameter tuning of the elastic net. *J. Optim. Theory Appl.* 183:993-1027, 2019.

[21] H. Calvete, C. Galé, On linear bilevel problems with multiple objectives at the lower level. *Omega.* 39:33-40, 2011.

[22] M. Breton, A. Alj, A. Haurie. Sequential Stackelberg equilibrium in two-person games. *J. Optim. Theory Appl.* 59:71-97, 1988.

[23] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Springer-Verlag, London 2013.

[24] B. Colson, P. Marcotte, G. Savard. An overview of bilevel optimization. *Ann. Op. Research.* 153:235-256, 2007.

[25] B.D. Craven. Aspects of multicriteria optimization, In *Recent Prospects in Mathematical Programming.* Gordon and Breach, 1991.

[26] J. P. Dauer. Optimization over the efficient set using an active constraint approach.*Z. Oper. Res.* 35:185-195, 1991.

[27] J.P. Dauer and T.A. Fosnaugh, Optimization over the efficient set, *J. Global Optim.* 7:261-277, 1995.

[28] S. Dempe, P. Mehlitz. Semivectorial bilevel programming versus scalar bilevel programming. *Optimization.* 69:657-679, 2020.

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

[30] S. Dempe, Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints. *Optimization.* 52:333-359, 2003.

[31] S. Dempe, N. Gadhi, A.B. Zemkoho. New optimality conditions for the semivectorial bilevel optimization problem. *J. Optim. Theory Appl.* 157:54-74, 2013.

[32] F. Y. Edgeworth. *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences.* C.K. Paul & co., London, 1881.

[33] G. Eichfelder. Multiobjective bilevel optimization. *Math. Program. Ser. A.* 123:419-449, 2010.

[34] J. Fülöp, A cutting plane algorithm for linear optimization over the efficient set. In *Generalized Convexity.* Lecture notes in Economics and Mathematical System 405, Springer-Verlag, 1994.

[35] A. M. Geoffrion. Proper efficiency and the theory of vector maximization. *J. Math. Anal. Appl.* 22:618-630, 1968.

[36] R. B. Holmes. *Geometric functional analysis and its applications*, Springer, New York, 1975.

[37] R. Horst and N.V. Thoai, Maximizing a concave function over the efficient or weakly-efficient set, European J. Oper. Res., 117, 1999, 239-252.

[38] R. Horst, N.V. Thoai, Y. Yamamoto, D. Zenke. On Optimization over the Efficient Set in Linear Multicriteria Programming. *J. Optim. Theory Appl.,*134:433-443, 2007.

[39] J. Jahn. *Vector Optimization.* Springer-Verlag, Berlin, Heidelberg, New York, 2004.

[40] H. W. Kuhn and A. W. Tucker. Nonlinear Programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, Berkley, 1951.

[41] D. T. Luc. *Theory of Vector Optimization.* Lecture Notes in Econom. and Math. Systems 319, Springer-Verlag, Berlin, Heidelberg, 1989.

[42] M. B. Lignola and J. Morgan, Stability of regularized bilevel programming problems, J. Optim. Theory Appl. 93, (1997) pp. 575-596.

[43] P. Loridan and J. Morgan, New results on approximate solutions in two-level optimization, Optimization 20, (1989), pp. 819-836.

[44] K. M. Miettinen. *Nonlinear Multiobjective Optimization.* Kluwer Academic Publishers, 1998.

[45] J. Morgan. Constrained well-posed two-level optimization problems. In *Nonsmooth Optimization and related topics.* (F. Clarke, V. F. Dem'yanov, and F. Giannessi, eds.), Ettore Majorana International Sciences Series, Plenum Press, New York, 1989.

[46] V. Pareto. *Cours d'économie politique.* F. Rouge, Lausanne, 1896.

[47] J. Philip. Algorithms for the vector maximization problem. *Math. Program.* 2:207-229, 1972.

[48] A. Ren, Y. Wang. A novel penalty function method for semivectorial bilevel programming problem. *Applied Mathematical Modelling* 40:135-149, 2016.

[49] H.v. Stackelberg, *Marktform und Gleichgewicht, Springer-Verlag, Berlin, (1934).* engl. transl.:*The Theory of the Market Economy,* Oxford University Press, 1952.

[50] Y.Zheng, Z. Wan. A solution method for semivectorial bilevel programming problem via penalty method. *J. Appl. Math. Comput.* 37:207-219, 2011.

[51] Y. Yamamoto. Optimization over the efficient set : overview. *J. Global Optim.* 22:285-317, 2002.

[52] P.L. Yu. *Multiple-Criteria Decision Making.* Plenum Press, New York and London, 1985.