

# An external penalty-type method for multicriteria\*

Ellen H. Fukuda<sup>†</sup>    L. M. Graña Drummond<sup>‡</sup>    Fernanda M. P. Raupp<sup>§</sup>

November 30, 2015

## Abstract

We propose an extension of the classical real-valued external penalty method to the multicriteria optimization setting. As its single objective counterpart, it also requires an external penalty function for the constraint set, as well as an exogenous divergent sequence of nonnegative real numbers, the so-called penalty parameters, but, differently from the scalar procedure, the vector-valued method uses an auxiliary function, which can be chosen among large classes of “monotonic” real-valued mappings. We analyze the properties of the auxiliary functions in those classes and exhibit some examples. The convergence results are similar to those of the scalar-valued method and, depending on the kind of auxiliary function used in the implementation, under standard assumptions, the generated infeasible sequences converge to weak Pareto or Pareto optimal points. We also propose an implementable local version of the external penalization method and study its convergence results.

**Keywords:** Constrained multiobjective optimization, external penalty method, Pareto optimality, scalar representation.

**AMS subject classifications:** 90C29, 90C30.

## 1 Introduction

Constrained multicriteria minimization problems appear frequently in many different areas, such as statistics [2], management science [12, 19], environmental analysis [14], space exploration [17], design [6] and engineering [3]. There are many strategies for solving such problems; one of the most popular techniques is the weighting method, where one minimizes a linear combination of the objectives. Its main drawback is the fact that we do not know

---

\*This work was supported by Grant-in-Aid for Young Scientists (B) (26730012) from Japan Society for the Promotion of Science, and a Grant (311165/2013-3) from National Counsel of Technological and Scientific Development (CNPq).

<sup>†</sup>Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan ([ellen@i.kyoto-u.ac.jp](mailto:ellen@i.kyoto-u.ac.jp)).

<sup>‡</sup>Faculty of Business and Administration, Federal University of Rio de Janeiro, Rio de Janeiro 22290-240, Brazil ([bolsigeno@gmail.com](mailto:bolsigeno@gmail.com)).

<sup>§</sup>National Laboratory for Scientific Computing, Rio de Janeiro 25651-075, Brazil ([fernanda@lncc.br](mailto:fernanda@lncc.br)).

a priori which are the suitable *weights* of this combination, i.e., those that do not lead to unbounded problems. Some extensions of classical real-valued methods to the vector-valued setting have been recently proposed to overcome that disadvantage. In this work we propose an extension of Zangwill’s external penalty method [20], which is a well-known technique for constrained scalar optimization problems. It is an iterative real-valued method that consists in adding a certain term to the objective function, in such a way that its value increases with the violation of the original restrictions, and then minimizing this penalized function in the whole space. In general, solutions of the penalized problems are infeasible, but, for large values of the so-called penalty parameters, the iterates are close to the constraint set. In other words, one substitutes a constrained problem by a sequence of unconstrained ones, for which we do have efficient solving techniques. Under reasonable assumptions, the sequence of those unconstrained minimizers converges to an optimal point of the original problem.

As far as we know, there was one attempt to generalize this strategy to multiobjective optimization: in 1984, White proposed a method [18] that, at each iteration, requires the Pareto unconstrained minimization of the penalized objective. In order to obtain Pareto optimality of the accumulation points of the generated sequences, an extra condition (not necessary in the scalar case) is required.

Here, we present a vector-valued version of Zangwill’s procedure, which, at each iteration, requires solving an unconstrained scalar problem. This new extension shares some features with its real-valued counterpart. For instance, in general, the generated sequence is also infeasible and, along it, the penalized objective values are nondecreasing. Moreover, as in the real-valued case, all accumulation points, if any, are optima of the original problem. Finally, the conditions under which the sequence fully converges to an optimal point are generalizations of the hypotheses required in the scalar case, and no additional assumptions are needed. Besides the (diverging) parameter sequence of positive real numbers and the penalty function, the proposed method uses an auxiliary function (whose presence in the scalar case would be irrelevant) and, depending on which one is chosen, the convergence will be to Pareto or weak Pareto optimal points.

The strategy proposed here has similar properties as other classical scalar-valued methods’ extensions, as the steepest descent [5, 11], the projected gradient [7, 8, 9], Newton [4, 10] and the proximal point [1] vector-valued methods. All these extensions have in common an important feature: the iterates are computed by solving scalar optimization subproblems; moreover, each iterate could be implicitly obtained by the application of the corresponding real-valued method to a certain (a priori unknown) linear combination of the objectives. All of these procedures just seek a single Pareto (or weak Pareto) point and the convergence results are natural extensions of their scalar correlatives. Nevertheless, as shown in [4], through numerical experiments, by initializing them with randomly chosen points, in some cases we can expect to obtain good approximations of the optimal set.

The outline of this paper is as follows. In Section 2, we introduce the problem, as well as the notion of vector external penalty function with some examples. In Section 3, we define a couple of auxiliary functions classes, state their properties, and show examples. In Section 4, we present the external penalty method for multiobjective optimization; we make

some comments, study its behavior and exhibit a very simple example, in which it works far better than the weighting method. The convergence analysis is in Section 5; basically, it consists of the extension of the classical results for the single objective case. We also propose an implementable local version of the method, analyzing its convergence. Additionally, a brief comparison with White's method is presented. In Section 6, we present another family of auxiliary functions and show an example in which, by simply varying a single parameter on the auxiliary function, we can retrieve the whole optimal set. Finally, in Section 7, we make some final remarks, briefly commenting on when the whole (weak) Pareto frontier can be obtained by using the method.

## 2 The problem and the penalty-type functions

Let  $\mathbb{R}^m$  be endowed with the partial order induced by  $\mathbb{R}_+^m$ , the nonnegative orthant of  $\mathbb{R}^m$ , given by

$$u \leq v \quad \text{if} \quad u_j \leq v_j \quad \text{for all} \quad j = 1, \dots, m.$$

We also consider the following stronger relation defined in  $\mathbb{R}^m$ :

$$u < v \quad \text{if} \quad u_j < v_j \quad \text{for all} \quad j = 1, \dots, m.$$

Given a continuous function  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  and a nonempty closed set  $D \subseteq \mathbb{R}^n$ , consider the following constrained vector-valued optimization problem:

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && x \in D, \end{aligned} \tag{1}$$

understood in the Pareto or weak Pareto sense.

We recall that a point  $x^* \in D$  is a *weak Pareto optimal* solution of (1) if there does not exist  $x \in D$  such that  $f(x) < f(x^*)$ . The point  $x^* \in D$  is a *Pareto optimal* solution of (1) if there does not exist  $x \in D$  such that  $f(x) \leq f(x^*)$  and  $f_{j_0}(x) < f_{j_0}(x^*)$  for at least one  $j_0 \in \{1, \dots, m\}$ . Note that a Pareto optimal solution is a weak Pareto optimum.

We intend to find Pareto or weak Pareto solutions of the above problem by means of an external penalty-type method. For the vector-valued case, we may define a penalty function for the feasible set as follows.

**Definition 2.1.** *A vector external penalty function for the set  $D$  is a continuous function  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$  such that  $P(x) = 0$  if and only if  $x \in D$ .*

Note that, when  $m = 1$ , we retrieve the classical notion of external penalty function used in scalar-valued optimization [16]. In order to ensure that  $\tilde{x} \in D$ , this point must satisfy the following system of nonlinear equations

$$P_j(\tilde{x}) = 0, \quad j = 1, \dots, m.$$

But, whenever each  $P_j$  is a (scalar-valued) external penalty function for  $D$ , then, for guaranteeing that  $\tilde{x} \in D$ , it suffices that  $P_j(\tilde{x}) = 0$  for just one  $j$  (no matter which one). A

particular case of this situation is when  $P_j = \hat{P}$  for all  $j$ , where  $\hat{P}$  is a scalar-valued external penalty function for  $D$ .

Differently from the scalar case, a vector external penalty function  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$  does not necessarily satisfy  $P(x) > 0$  if and only if  $x \notin D$ . Indeed, for  $m \geq 2$ , it suffices to take, for instance,  $P_1(x) = 0$  for all  $x \in \mathbb{R}^n$  and, for  $j \neq 1$ ,  $P_j$  such that  $P$  is a vector external penalty for  $D$ . But, if we take  $P$  such that, for all  $i$ ,  $P_i$  is a (scalar-valued) external penalty function for  $D$ , then we clearly do have that  $P(x) > 0$  if and only if  $x \notin D$ ; in particular, this happens to those vector external penalty functions that have all components equal to a certain real-valued external penalty function for  $D$ .

## 2.1 Examples of external penalty functions

We now give some examples of external penalty-type functions.

**Example 2.2.** Consider the following classical case of a constraint set for both scalar and vector-valued optimization:

$$D := \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}, \quad (2)$$

where  $g: \mathbb{R}^n \rightarrow \mathbb{R}^q$  and  $h: \mathbb{R}^n \rightarrow \mathbb{R}^r$  are continuous functions. For simplicity, let us suppose that  $r + q \leq m$ . Take

$$p(x)^\top := \left( \max\{0, g_1(x)\}, \dots, \max\{0, g_q(x)\}, h_1(x), \dots, h_r(x) \right)$$

for all  $x \in \mathbb{R}^n$ , with  $\top$  denoting the transpose. Define  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$  as follows:

$$\begin{aligned} P_i(x) &= |p_i(x)|^\beta, & i = 1, \dots, r + q, \\ P_i(x) &= 0, & i = r + q + 1, \dots, m \end{aligned}$$

for all  $x \in \mathbb{R}^n$ , where  $\beta \geq 1$ . Since  $P$  is continuous and  $P(x) = 0$  if and only if  $x \in D$ ,  $P$  is an external penalty function for  $D$ . As we know, if  $f, g, h$  are smooth,  $P$  is also differentiable for  $\beta > 1$ . If  $r + q > m$ , we can simply take  $P(x)^\top = (\sum_{i=1}^{r+q} |p_i(x)|^\beta, 0, \dots, 0)$  or  $P(x) = \sum_{i=1}^{r+q} |p_i(x)|^\beta e$ , where  $e = (1, \dots, 1)^\top \in \mathbb{R}^m$ .

In the next example we show that, as in some scalar problems, the distance from a point to a constraint set  $D$  can be taken as an external penalty function. First, let us mention that, from now on,  $\|\cdot\|$  will always stand for the Euclidean norm, i.e.,  $\|x\|^2 := \langle x, x \rangle$ , where  $\langle x, y \rangle := \sum_{i=1}^n x_i y_i$  for all  $x, y \in \mathbb{R}^n$ .

**Example 2.3.** Let us consider the case in which  $D$  is defined by finitely many homogeneous linear equations. Assume that  $D := \bigcap_{i=1}^r D_i$ , where  $D_i \subset \mathbb{R}^n$  is a hyperplane, say the orthogonal of a norm one vector  $w^i \in \mathbb{R}^n$  for all  $i = 1, \dots, r$ , with  $r \leq m$ . We can take  $P_i(x) = 0$  for  $i = r + 1, \dots, m$  and  $P_i(x)$  as the distance between  $x$  and the  $(n - 1)$ -dimensional subspace  $D_i$ , i.e.,  $P_i(x) = \|x - \Pi_{D_i}(x)\|$  for  $i = 1, \dots, r$ , where  $\Pi_{D_i}: \mathbb{R}^n \rightarrow \mathbb{R}^n$  is the orthogonal projector onto  $D_i$ , i.e.,  $\Pi_{D_i}(x) := x - \langle w^i, x \rangle w^i$ , so

$$\begin{aligned} P_i(x) &= |\langle w^i, x \rangle|, & i = 1, \dots, r, \\ P_i(x) &= 0, & i = r + 1, \dots, m. \end{aligned}$$

In particular, if we have just one hyperplane, say the orthogonal of a norm one vector  $w \in \mathbb{R}^n$ , we can take  $P(x) = |\langle w, x \rangle|e$ , with  $e = (1, \dots, 1)^\top$ .

In general, if  $D_i$  is a closed convex nonempty set of  $\mathbb{R}^n$  for  $i = 1, \dots, r$ , with  $r \leq m$ , we can take the following external penalty function for  $D$ :

$$\begin{aligned} P_i(x) &= \text{dist}(x, D_i), & i &= 1, \dots, r, \\ P_i(x) &= 0, & i &= r + 1, \dots, m, \end{aligned}$$

where, as usual,  $\text{dist}(x, D_i) := \min_{y \in D_i} \|x - y\|$ .

In the case that  $r > m$ , we can take  $P(x)^\top := (\sum_{i=1}^m \text{dist}(x, D_i), 0, \dots, 0)$  or, simply,  $P(x) := \sum_{i=1}^m \text{dist}(x, D_i)e$ , where, once again,  $e$  stands for the  $m$ -vector of ones.

### 3 Auxiliary functions

In this section we introduce some auxiliary functions necessary for the vector-valued external penalty method. First, we consider a generalization of the notion of increasing function for the vector-valued case, then we present some examples.

#### 3.1 Assumptions on auxiliary functions

We now present different notions of monotonic vector-valued functions. We point out that these concepts are not new. For a more general setting, we can find them in [13, Chapter 5, Definition 5.1 (b), (c)], as well as in [15, Chapter 1, Definition 4.1], where even more general cases are considered.

**Definition 3.1.** A function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  is called strictly (monotonically) increasing or weakly-increasing ( $w$ -increasing) if, for all  $u, v \in \mathbb{R}^m$ ,

$$u < v \quad \Rightarrow \quad \Phi(u) < \Phi(v),$$

and  $\Phi$  is called strongly (monotonically) increasing ( $s$ -increasing) if for all  $u, v \in \mathbb{R}^m$ ,

$$u \leq v \text{ and } u_{j_0} < v_{j_0} \text{ for at least one } j_0 \quad \Rightarrow \quad \Phi(u) < \Phi(v).$$

Clearly, an  $s$ -increasing function is also  $w$ -increasing. A simple example of a  $w$ -increasing function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$ , which is not  $s$ -increasing, is given by  $\Phi(u) := \max_{i=1, \dots, m} \{u_i\}$ . A general example of  $w$ -increasing function which is not  $s$ -increasing is the following:  $\Phi(u) := \sum_{i=1}^m \phi_i(u_i)$ , where  $\phi_i: \mathbb{R} \rightarrow \mathbb{R}$  is nondecreasing for  $i = 1, \dots, m$  and there exists at least one index  $i_0$  such that  $\phi_{i_0}$  is increasing and not all  $\phi_i$ 's are increasing functions. Moreover, an example of an  $s$ -increasing (bounded) function is  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$ , given by  $\Phi(u) := \sum_{i=1}^m \arctan(u_i)$ . More generally,  $\Phi(u) := \sum_{i=1}^m \phi_i(u_i)$ , where  $\phi_i: \mathbb{R} \rightarrow \mathbb{R}$  is increasing for  $i = 1, \dots, m$ , is an  $s$ -increasing function (not necessarily bounded).

It is easy to see that if the function  $\Phi$  is  $w$ -increasing and continuous, then we have

$$u \leq v \quad \Rightarrow \quad \Phi(u) \leq \Phi(v) \quad \text{for any } u, v \in \mathbb{R}^m. \quad (3)$$

For the sake of completeness, we state a simple result for monotonic functions which relates optimality for the scalar-valued problem  $\min_{x \in D} \Phi(f(x))$  to weak Pareto and Pareto optimality for the vector-valued problem (1).

**Lemma 3.2.** *1. If  $\Phi$  is a  $w$ -increasing function and  $x^* \in \operatorname{argmin}_{x \in D} \Phi(f(x))$ , then  $x^*$  is a weak Pareto optimal solution for problem (1).*

*2. If  $\Phi$  is an  $s$ -increasing function and  $x^* \in \operatorname{argmin}_{x \in D} \Phi(f(x))$ , then  $x^*$  is a Pareto optimal solution for problem (1).*

*Proof.* The results follow from [13, Chapter 5, Lemmas 5.14 and 5.24]. □

For the definition of the method, we need to introduce two classes of auxiliary functions. We begin with those which are continuous, monotonic and overestimate the canonical projections.

**Definition 3.3.** *A continuous  $w$ -increasing function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  is of weak type (or  $w$ -type) if it satisfies the following property:*

$$u_j \leq \Phi(u) \quad \text{for any } u \in \mathbb{R}^m \text{ and all } j = 1, \dots, m. \quad (\mathcal{P})$$

*Similarly, a continuous  $s$ -increasing function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  is of strong type (or  $s$ -type) if it satisfies  $(\mathcal{P})$ .*

Clearly, an  $s$ -type function is necessarily of  $w$ -type, because  $s$ -increasing functions are  $w$ -increasing. Let us now introduce the other class of auxiliary functions which will also be used in the method we are about to present.

**Definition 3.4.** *A function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  continuous,  $w$ -increasing and subadditive (i.e.,  $\Phi(u + v) \leq \Phi(u) + \Phi(v)$  for all  $u, v \in \mathbb{R}^m$ ) is weakly-subadditive (or  $w$ -subadditive), if given a sequence  $\{u^k\} \subset \mathbb{R}_+^m$  and  $M > 0$  there exists  $T > 0$  such that*

$$\Phi(u^k) \leq M \text{ for all } k \implies \|u^k\| \leq T \text{ for all } k. \quad (\mathcal{Q})$$

*If  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  is continuous,  $s$ -increasing, subadditive, and satisfies property  $(\mathcal{Q})$ , it is a strongly-subadditive (or  $s$ -subadditive) function.*

Property  $(\mathcal{Q})$  tells us that there is no unbounded set in  $\mathbb{R}_+^m$  with bounded image via  $\Phi$ . Note also that, even though we now ask a stronger property on the auxiliary function  $\Phi$ , namely, its subadditivity, we relax property  $(\mathcal{P})$ . Indeed, property  $(\mathcal{Q})$  is weaker than  $(\mathcal{P})$ , since any function  $\Phi$  that satisfies property  $(\mathcal{P})$  also verifies  $(\mathcal{Q})$  with  $T = \sqrt{m}M$ .

### 3.2 Examples of auxiliary functions

First, let us exhibit some general examples of  $w$ -type and  $s$ -type functions.

**Example 3.5.** Let  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  be defined by

$$\Phi(u) := \max_{i=1, \dots, m} \{u_i + a_i\} + b,$$

with  $a_i + b \geq 0$  for all  $i = 1, \dots, m$ . Then,  $\Phi$  is a  $w$ -type function but not an  $s$ -type one. In particular,  $\Phi(u) := \max_{i=1, \dots, m} \{u_i\}$  is also a  $w$ -type function but not an  $s$ -type one.

**Example 3.6.** Let  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  be given by

$$\Phi(u) := \max_{i=1, \dots, m} \{\phi_i(u) + \xi_i(u)\} + \zeta(u),$$

where  $\phi_i, \xi_i, \zeta: \mathbb{R}^m \rightarrow \mathbb{R}$  are continuous functions, with  $\xi_i(u) + \zeta(u) \geq 0$ , such that  $\phi_i$  satisfies  $(\mathcal{P})$  for all  $i$  and  $\Phi(u)$  is  $w$ -increasing (e.g., if  $\phi_i, \xi_i$  and  $\zeta$  are  $w$ -increasing for all  $i$ ). Then  $\Phi$  is a  $w$ -type function, but not necessarily an  $s$ -type function. Note that, if one of these three functions is  $s$ -increasing for all  $i = 1, \dots, m$  and the other two are  $w$ -increasing, then  $\Phi$  is an  $s$ -type function.

**Example 3.7.** Let  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  be a function defined by

$$\Phi(u) := \psi_1(u_1) + \dots + \psi_m(u_m),$$

where  $\psi_i: \mathbb{R} \rightarrow \mathbb{R}_+$  is a continuous increasing function satisfying property  $(\mathcal{P})$  for all  $i = 1, \dots, m$ , i.e.,  $t \leq \psi_i(t)$  for all  $t \in \mathbb{R}$  and all  $i$  (e.g.,  $\psi_i(u_i) := a_i \exp(u_i)$ , with  $a_i \geq 1$ ). Then,  $\Phi$  is an  $s$ -type function.

**Example 3.8.** If  $\Phi_1, \dots, \Phi_r: \mathbb{R}^m \rightarrow \mathbb{R}$  are  $w$ -type ( $s$ -type) functions and  $\alpha_1, \dots, \alpha_r$  are nonnegative scalars adding up 1, then

$$\Phi := \sum_{i=1}^r \alpha_i \Phi_i$$

is also of weak type (strong type). Clearly, linear combinations of nonnegative  $w$ -type ( $s$ -type) functions with all scalars greater than or equal to 1 are also of the same type.

**Example 3.9.** Assume that  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  is a  $w$ -type function. Let  $\omega \in \mathbb{R}^m$  and define  $\hat{\omega} := \max_{i=1, \dots, m} |\omega_i|$  or  $\hat{\omega} := a \exp(|\omega_1| + \dots + |\omega_m|)$ , where  $a \geq 1$ . Then, the function  $\Phi_\omega: \mathbb{R}^m \rightarrow \mathbb{R}$ , defined by

$$\Phi_\omega(u) := \Phi(u + \omega) + \hat{\omega}$$

is of  $w$ -type. Moreover, if  $\Phi$  is an  $s$ -type function, then  $\Phi_\omega$  is of  $s$ -type.

**Example 3.10.** Let  $\Psi: \mathbb{R}^m \rightarrow \mathbb{R}$  be an  $s$ -type function,  $\Phi, \Upsilon: \mathbb{R}^m \rightarrow \mathbb{R}$   $w$ -type functions, and  $e := (1, \dots, 1)^\top \in \mathbb{R}^m$ . Then,

$$u \mapsto \Phi(\Upsilon(u)e) \quad \text{and} \quad u \mapsto \Phi(\Psi(u)e)$$

are  $w$ -type and  $s$ -type functions, respectively. If  $\psi: \mathbb{R} \rightarrow \mathbb{R}$  is a continuous increasing function satisfying property  $(\mathcal{P})$ , then the compositions of functions

$$\psi \circ \Psi \quad \text{and} \quad \psi \circ \Phi$$

are  $s$ -type and  $w$ -type functions, respectively.

Besides the max-type ones, the previous examples of  $w$ -type auxiliary functions are basically compositions of inner products with continuous increasing scalar-valued functions. Next example shows that these are not all the possibilities.

**Example 3.11.** Let  $\Psi: \mathbb{R}^m \rightarrow \mathbb{R}$  be defined by

$$\Psi(u) = \Phi(u)\Upsilon(u),$$

where  $\Phi$  is as in Example 3.7 and  $\Upsilon: \mathbb{R}^m \rightarrow \mathbb{R}$  is a continuous  $w$ -increasing function such that  $\Upsilon(u) \geq 1$  for all  $u$  (e.g.  $\Upsilon(u) = \sum_{i=1}^m \gamma_i(u_i)$ , with  $\gamma_i(t) = \arctan(t) + \pi$  for all  $i$ ). Then,  $\Psi$  is an  $s$ -type function.

Now, we exhibit some examples of weakly and strongly subadditive auxiliary functions.

**Example 3.12.** Consider  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$ , defined by  $\Phi(u) = \max_{i=1, \dots, m} \{u_i\}$ . Clearly,  $\Phi$  is subadditive and, as we saw in Example 3.5, it is of  $w$ -type, so  $\Phi$  is a  $w$ -subadditive function. And, since  $\Phi$  is not  $s$ -increasing, it is not an  $s$ -subadditive function.

**Example 3.13.** Take  $a \in \mathbb{R}^m$  such that  $a > 0$  and let the linear function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  be defined by

$$\Phi(u) = \sum_{i=1}^m a_i u_i.$$

For any  $u \in \mathbb{R}^m$  such that  $u \geq 0$ , we have that  $\Phi(u) \geq \kappa \sum_{i=1}^m u_i = \kappa \sum_{i=1}^m |u_i| =: \kappa \|u\|_1$ , where  $\kappa := \min_{i=1, \dots, m} a_i > 0$ . So, since  $\|u\| \leq \|u\|_1$ , property (Q) holds with  $T = M/\kappa$ . As  $\Phi$  is  $s$ -increasing and satisfies (Q), it is an  $s$ -subadditive function. Observe that  $\Phi$  is not an  $s$ -type function, since  $\Phi(\frac{1}{a_1}, \frac{-1}{a_2}, 0, \dots, 0) = 0 \not\geq \frac{1}{a_1}$ , and so property (P) does not hold.

**Example 3.14.** Let  $\Psi, \Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  be  $w$ -subadditive ( $s$ -subadditive) functions such that  $\Phi(x) \geq 0$  for all  $x \in \mathbb{R}^m$ . Then the composition  $x \mapsto \Psi(\Phi(x)e)$ , where  $e$  is the  $m$ -vector of ones, is also  $w$ -subadditive ( $s$ -subadditive).

As we will see in the next section, any of these  $w$ -type ( $s$ -type),  $w$ -subadditive ( $s$ -subadditive) functions can be employed in the algorithm. Properties (P) and (Q) will be used in the convergence proofs. Nevertheless, in practical terms, we do not always need them. Indeed, let us recall Example 3.9, and note that the minimizers of  $x \mapsto \Phi_\omega(f(x) + \rho_k P(x)) = \Phi(f(x) + \rho_k P(x) + \omega) + \hat{\omega}$  are the same as those of  $x \mapsto \Psi_\omega(f(x) + \rho_k P(x))$ , where  $\Psi_\omega := \Phi_\omega - \hat{\omega}$  is a continuous  $w$ -increasing function which does not satisfy property (P) and it is not subadditive, so it is neither of  $w$ -type nor  $w$ -subadditive. But  $\Psi_\omega$  can be used to generate the same sequence of iterates (and, therefore, with the same convergence properties) as the one produced by the auxiliary function  $\Phi_\omega$ .

## 4 An external penalty-type method for multiobjective optimization

In this section, we define the *multicriteria external penalty method (MEPM)* for solving problem (1). First, let us consider the weak version of the method. Let  $\mathbb{R}_{++}$  be the



set of positive real numbers. Take  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$  a vector external penalty for  $D \subseteq \mathbb{R}^n$ ,  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$  a  $w$ -type or  $w$ -subadditive function and  $\{\rho_k\} \subset \mathbb{R}_{++}$  a divergent sequence, such that  $\rho_{k+1} > \rho_k$  for all  $k$ . The method is iterative and generates a sequence  $\{x^k\} \subset \mathbb{R}^n$  by

$$x^k \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \Phi\left(f(x) + \rho_k P(x)\right), \quad k = 1, 2, \dots \quad (4)$$

The strong version of the method is formally identical to the weak one, but with  $\Phi$  as an  $s$ -type or  $s$ -subadditive function.

Let us make some comments and observations concerning both versions of the method.

1. Note that, as in the classical real-valued external penalty method, some hypotheses are needed in order to guarantee the existence of  $x^k$  for all  $k$ . For instance, we can apply the method for any functions  $f$ ,  $\Phi$  and  $P$  such that  $x \mapsto \Phi(f(x) + \rho_k P(x))$  is coercive in  $\mathbb{R}^n$  for all  $k$ . In this case, by continuity, we have  $\underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \Phi(f(x) + \rho_k P(x)) \neq \emptyset$ .
2. In both versions of *MEPM*, a necessary condition for the welldefinedness of the whole sequence  $\{x^k\}$  is the following:

$$-\infty < \inf_{x \in D} \Phi(f(x)). \quad (5)$$

Indeed, for any  $\tilde{x} \in D$ , from the fact that  $P$  is a penalty function for  $D$ , we get

$$-\infty < \Phi(f(x^k) + \rho_k P(x^k)) = \min_{x \in \mathbb{R}^n} \Phi(f(x) + \rho_k P(x)) \leq \Phi(f(\tilde{x}) + \rho_k P(\tilde{x})) = \Phi(f(\tilde{x})).$$

Since  $\tilde{x}$  is an arbitrary element of  $D$ , condition (5) follows.

3. This method inherits some features and drawbacks of its real-valued counterpart. Firstly, we mention that it does not have any kind of “memory”, i.e., the former iterate is not used to compute the current one; nevertheless, in order to obtain  $x^k$ , it seems reasonable to initialize the subroutine used to (approximately) solve subproblem (4) with the former iterate  $x^{k-1}$ . Secondly, the benefit of applying *MEPM* is to change a constrained (vector-valued) problem by a sequence of unconstrained (scalar-valued) ones with continuous objective functions.
4. When  $m = 1$ , taking  $\Phi(u) = \max_{i=1, \dots, m} \{u_i\}$ , we retrieve the classical (scalar-valued) external penalty method. Actually, in the scalar case, any auxiliary function  $\Phi: \mathbb{R} \rightarrow \mathbb{R}$  is increasing, so iteration (4) generates the same sequence as Zangwill’s method.

One may ask why it is worth to use this method instead of others. We observe that problem (1) may have a very poor structure:  $f$  is just required to be continuous. Whenever  $\Phi(u) = \max_{i=1, \dots, m} \{u_i\}$  and  $P_j = \hat{P}$ , where  $\hat{P}: \mathbb{R}^n \rightarrow \mathbb{R}$  for all  $j$ , *MEPM* is just the scalar-valued external penalty method applied to the minimization of the continuous function  $x \mapsto \max_{i=1, \dots, m} \{f_i(x)\}$  in  $D$ , with  $\hat{P}$  as a penalty function.

5. One may ask why we should use this method instead of applying the classical scalar external penalty method to problem  $\min_{x \in D} \Phi(f(x))$ . An answer to this question is that, *MEPM* has more degrees of freedom: we do not always need to choose a max-type auxiliary function  $\Phi$  and nor do we have to use a penalty of the type  $P = (\hat{P}, \dots, \hat{P})$ , where  $\hat{P}$  is a scalar-valued penalty for  $D$ .
6. As mentioned in the introduction, *MEPM* shares the following feature with other extensions of classical scalar methods to the vectorial setting: under certain regularity conditions, all iterates are implicitly obtained by the application of the corresponding real-valued algorithm to a certain weighted scalarization. In order to see this assertion, assume that  $f$  and  $P = (\hat{P}, \dots, \hat{P})^\top$  are  $\mathbb{R}_+^m$ -convex (i.e.,  $f_j$  and  $\hat{P}$  are convex for all  $j$ ) differentiable functions, with  $\hat{P}$  a scalar-valued penalty for  $D$ . Let  $\Phi$  be defined by  $\Phi(u) = \max_{i=1, \dots, m} \{u_i\}$  for all  $u \in \mathbb{R}^m$ . Next, reformulate  $\min_{x \in \mathbb{R}^n} \max_{i=1, \dots, m} \{f_i(x) + \rho_k P_i(x)\}$  as the minimization of  $t$  subject to  $f_j(x) + \rho_k P_j(x) \leq t$ ,  $j = 1, \dots, m$ , a smooth problem in  $(x, t) \in \mathbb{R}^n \times \mathbb{R}$ . Now, from the first order optimality condition of the above reformulation, we see that

$$x^k \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \langle \lambda^k, f(x) \rangle + \rho_k \hat{P}(x),$$

which means that  $\{x^k\}$  can be obtained via the application of the classical (scalar) external penalty method to the real-valued function  $x \mapsto \langle \lambda^k, f(x) \rangle$ , a weighted scalarization of the vector-valued objective  $f$ , with weighting vector given by  $\lambda^k \in \mathbb{R}_+^m$ , using the scalar-valued penalty  $\hat{P}(x)$  for  $D$  and  $\{\rho_k\}$  as the parameter sequence. Of course we do not know, a priori, the nonnegative weights  $\lambda_1^k, \dots, \lambda_m^k$ , which add up one.

The next proposition establishes a simple condition under which both versions of *MEPM* converge to optimal points in its very first iteration.

**Proposition 4.1.** *Consider MEPM implemented with an external penalty function  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$ , a sequence of parameters  $\{\rho_k\} \subset \mathbb{R}_{++}$  and a  $w$ -type or  $w$ -subadditive function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$ . If we have*

$$\underset{x \in D}{\operatorname{argmin}} \Phi(f(x)) = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \Phi(f(x)),$$

*then the method converges in one iteration to a weak Pareto solution of problem (1). If  $\Phi$  is an  $s$ -type or  $s$ -subadditive function, MEPM converges in a single iteration to a Pareto optimum of (1).*

*Proof.* If  $x^* \in \underset{x \in D}{\operatorname{argmin}} \Phi(f(x))$ , then  $P(x^*) = 0$  and so, combining the optimality of  $x^*$  in  $\mathbb{R}^n$ , the facts that  $P(x) \geq 0$  for all  $x \in \mathbb{R}^n$  and  $\rho_1 > 0$  with (3), we get

$$\Phi(f(x^*) + \rho_1 P(x^*)) = \Phi(f(x^*)) \leq \Phi(f(x)) \leq \Phi(f(x) + \rho_1 P(x)) \quad \text{for all } x \in \mathbb{R}^n.$$

Therefore,  $x^* \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \Phi(f(x) + \rho_1 P(x))$ , and so  $\Phi(f(x^1) + \rho_1 P(x^1)) = \Phi(f(x^*) + \rho_1 P(x^*)) = \Phi(f(x^*))$ . Hence, once again by (3), we obtain

$$\Phi(f(x^1)) \leq \Phi(f(x^1) + \rho_1 P(x^1)) = \Phi(f(x^*)) \leq \Phi(f(x)) \quad \text{for all } x \in \mathbb{R}^n.$$

Whence,  $x^1 \in \operatorname{argmin}_{x \in \mathbb{R}^n} \Phi(f(x)) = \operatorname{argmin}_{x \in D} \Phi(f(x))$ . The result then follows from Lemma 3.2. The strong result is also a consequence of Lemma 3.2.  $\square$

Let us show a very simple application of the above proposition.

**Example 4.2.** Consider  $n = 1$ ,  $m = 2$ ,  $D = [-1, +\infty)$  and  $f: \mathbb{R} \rightarrow \mathbb{R}^2$ , given by  $f(t) = (t, -\ell t)^\top$ , where  $\ell = 1, 2, \dots$ . In order to apply MEPM to this problem, we take  $\Phi(u) = \max_{i=1, \dots, m} \{u_i\}$ , a penalty function  $P$ , and  $\{\rho_k\}$  an increasingly divergent sequence of positive real numbers. It is easy to see that the condition required in Proposition 4.1 holds, and so MEPM converges in its first iteration to a Pareto point. We point out that the weighting method with scalarization parameter  $\alpha \in [0, 1]$  applied to this problem fails when  $\alpha \in [0, \ell/(1 + \ell)]$ , which means that for large  $\ell$ , it fails in a large set of weights.

We now show some elementary properties of sequences generated by MEPM which will be needed in the sequel. Since  $s$ -increasing functions are  $w$ -increasing, we just prove them for the weak version of the method.

**Lemma 4.3.** Let  $\tilde{x} \in D \subseteq \mathbb{R}^n$  and  $\{x^k\} \subset \mathbb{R}^n$  be a sequence generated by the weak version of MEPM implemented with a penalty function  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$ , a parameters sequence  $\{\rho_k\} \subset \mathbb{R}_{++}$  and a  $w$ -type or  $w$ -subadditive function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$ . Then, for all  $k = 1, 2, \dots$ , the following statements hold.

1. For any auxiliary function  $\Phi$ , we have

$$\Phi(f(x^k) + \rho_k P(x^k)) \leq \Phi(f(x^{k+1}) + \rho_{k+1} P(x^{k+1})) \leq \Phi(f(\tilde{x})) \quad \text{for all } j = 1, \dots, m.$$

If  $\Phi$  is of  $w$ -type, then we also have

$$f_j(x^k) + \rho_k P_j(x^k) \leq \Phi(f(x^k) + \rho_k P(x^k)) \quad \text{for all } j = 1, \dots, m. \quad (6)$$

2. For any auxiliary function  $\Phi$ , we have  $\Phi(f(x^k)) \leq \Phi(f(\tilde{x}))$ . If  $\Phi$  is of  $w$ -type, then we also have that  $f_j(x^k) \leq \Phi(f(\tilde{x}))$  for all  $j = 1, \dots, m$ .

3. For any auxiliary function  $\Phi$ , there exists  $\eta \in \mathbb{R}$  such that

$$\lim_{k \rightarrow \infty} \Phi(f(x^k) + \rho_k P(x^k)) = \eta.$$

*Proof.* 1. Using the properties of  $\Phi$  and the definitions of  $x^k$  and  $x^{k+1}$ , we obtain

$$\begin{aligned} \Phi(f(x^k) + \rho_k P(x^k)) &= \min_{x \in \mathbb{R}^n} \Phi(f(x) + \rho_k P(x)) \\ &\leq \Phi(f(x^{k+1}) + \rho_k P(x^{k+1})) \\ &\leq \Phi(f(x^{k+1}) + \rho_{k+1} P(x^{k+1})) \\ &= \min_{x \in \mathbb{R}^n} \Phi(f(x) + \rho_{k+1} P(x)) \\ &\leq \Phi(f(\tilde{x}) + \rho_{k+1} P(\tilde{x})) \\ &= \Phi(f(\tilde{x})), \end{aligned}$$

where the first equality follows from (4), the second inequality is a consequence of the weak monotonic behavior of  $\Phi$  combined with the facts that  $0 < \rho_k < \rho_{k+1}$  for all  $k$  and  $P(x) \geq 0$  for all  $x \in \mathbb{R}^n$ , and the last equality follows from the facts that  $P$  is a vector external penalty function for  $D$  and  $\tilde{x} \in D$ . If  $\Phi$  is of  $w$ -type, then (6) follows immediately from property  $(\mathcal{P})$ .

2. From the proof of item 1, we have that  $\Phi(f(x^k) + \rho_k P(x^k)) \leq \Phi(f(\tilde{x}))$ , so, from the facts that  $P \geq 0$ ,  $\rho_k > 0$  and  $\Phi$  is a  $w$ -increasing function, it follows that

$$\Phi(f(x^k)) \leq \Phi(f(\tilde{x})).$$

So, if  $\Phi$  is of  $w$ -type, from property  $(\mathcal{P})$ ,  $f_j(x^k) \leq \Phi(f(\tilde{x}))$  for all  $j = 1, \dots, m$ .

3. By item 1,  $\{\Phi(f(x^k) + \rho_k P(x^k))\}$  is a nondecreasing bounded real numbers sequence, so, as  $k \rightarrow \infty$ , it converges to some  $\eta \in \mathbb{R}$ . □

As in the classical real-valued method, from item 1 of the above lemma, we observe that in the vector-valued case we have

$$\Phi(f(x^k) + \rho_k P(x^k)) \leq \Phi(f(x^{k+1}) + \rho_{k+1} P(x^{k+1})) \quad k = 1, 2, \dots$$

for any sequence generated by *MEPM* implemented with a  $w$  or  $s$ -type,  $w$  or  $s$ -subadditive auxiliary function  $\Phi$ . We also know that in the scalar case, we have the following facts: the real sequences  $\{P(x^k)\}$  and  $\{f(x^k)\}$  are non-increasing and nondecreasing, respectively. However, in the general case ( $m \geq 2$ ), we may not have such properties. When we choose an arbitrary auxiliary function  $\Phi$ , even though we can not ensure that the functional values sequence is nondecreasing, we can, at least, say that they converge from below to the optimal values. Indeed, from item 2 of the last lemma, we have

$$\sup_{k=1,2,\dots} \Phi(f(x^k)) \leq \inf_{x \in D} \Phi(f(x)) \quad \text{and} \quad \sup_{k=1,2,\dots} f_j(x^k) \leq \inf_{x \in D} \Phi(f(x)) \quad \text{for all } j = 1, \dots, m.$$

## 5 Convergence analysis

Let us now study the convergence properties of sequences produced by both versions of *MEPM*. We begin with an extension of a classical result for real-valued optimization which establishes that accumulation points, if any, of a sequence generated by the external penalty method are optima of the original constrained minimization problem. We also show that, as in the real-valued method, whenever the sequence  $\{x^k\}$  has infinitely many iterates, all of them are infeasible points.

**Theorem 5.1.** *Let  $\{x^k\} \subset \mathbb{R}^n$  be a sequence generated by *MEPM* implemented with a penalty function  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$ , a parameters sequence  $\{\rho_k\} \subset \mathbb{R}_{++}$  and a  $w$ -type or  $w$ -subadditive function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$ .*

1. The point  $x^{k_0}$  belongs to  $D$  for some  $k_0$  if and only if  $x^{k_0}$  is a weak Pareto solution for problem (1).
2. If  $\bar{x}$  is an accumulation point of  $\{x^k\}$ , then  $\bar{x}$  is a weak Pareto optimum for the problem (1).

If *MEPM* is implemented with an  $s$ -type or an  $s$ -subadditive function  $\Phi$ , then items 1 and 2 hold with Pareto optimal solutions instead of weak Pareto optima.

*Proof.* 1. If  $x^{k_0}$  is a weak Pareto solution of (1), in particular,  $x^{k_0}$  belongs to  $D$ .

Conversely, let us now assume that  $x^{k_0} \in D$ . By item 2 of Lemma 4.3, we have

$$\Phi(f(x^{k_0})) \leq \inf_{x \in D} \Phi(f(x)),$$

and so  $x^{k_0}$  is a minimizer of  $\Phi(f(x))$  in  $D$ . Then, from Lemma 3.2,  $x^{k_0}$  is a weak-Pareto solution for problem (1).

2. Assume now that *MEPM* is implemented with a  $w$ -type function  $\Phi$ . Let  $K$  be an infinite subset of  $\{1, 2, \dots\}$  such that  $\lim_{K \ni k \rightarrow \infty} x^k = \bar{x}$ . Since  $f$  is a continuous function, we have that

$$\lim_{K \ni k \rightarrow \infty} f_j(x^k) = f_j(\bar{x}) \quad \text{for all } j = 1, \dots, m. \quad (7)$$

Therefore,

$$|f_j(x^k)| \leq M, \quad \text{for some } M > 0 \text{ for all } j = 1, 2, \dots, m \text{ and all } k \in K. \quad (8)$$

On the other hand, by item 1 of Lemma 4.3,

$$f_j(x^k) + \rho_k P_j(x^k) \leq \Phi(f(\tilde{x})) =: \tilde{f} \quad \text{for all } j = 1, 2, \dots, m \text{ and any } \tilde{x} \in D.$$

So, from (8) and the above inequality, for all  $j = 1, 2, \dots, m$  and all  $k \in K$ , we get

$$0 \leq \rho_k P_j(x^k) = [f_j(x^k) + \rho_k P_j(x^k)] - f_j(x^k) \leq \tilde{f} + M. \quad (9)$$

Therefore, since  $\rho_k \rightarrow +\infty$  and  $P_j(x^k) \geq 0$  for all  $j$  and  $k \in K$ , necessarily,

$$\lim_{K \ni k \rightarrow \infty} P_j(x^k) = 0 \quad \text{for all } j = 1, \dots, m$$

and, since all  $P_j$  are continuous,

$$P_j(\bar{x}) = 0 \quad \text{for all } j = 1, \dots, m.$$

Hence, we have

$$P(\bar{x}) = 0,$$

and so  $\bar{x} \in D$ .

Let us call  $\hat{v} := \inf_{x \in D} \Phi(f(x))$ , which is a real number, in view of (5). Applying item 2 of Lemma 4.3, we get

$$\Phi(f(x^k)) \leq \hat{v} \quad \text{for all } k = 1, 2, \dots$$

Letting  $K \ni k \rightarrow \infty$  in the above inequality, we obtain

$$\Phi(f(\bar{x})) \leq \hat{v}.$$

Using the fact that  $\bar{x} \in D$ , we conclude that  $\bar{x} \in \operatorname{argmin}_{x \in D} \Phi(f(x))$ , and the result follows from item 1 of Lemma 3.2.

Now let us suppose that *MEPM* is implemented with a  $w$ -subadditive function  $\Phi$ . Let  $K$  be an infinite subset of  $\{1, 2, \dots\}$  such that  $\lim_{K \ni k \rightarrow \infty} x^k = \bar{x}$ . From item 1 of Lemma 4.3,  $\Phi(f(x^k) + \rho_k P(x^k)) \leq \inf_{x \in D} \Phi(f(x)) =: \hat{v}$  for all  $k$  and so, due to the facts that  $\rho_1 \leq \rho_k$  for all  $k$ ,  $P$  is nonnegative and  $\Phi$  is  $w$ -increasing, we have

$$\Phi(f(x^k) + \rho_1 P(x^k)) \leq \hat{v} \quad \text{for all } k \in K. \quad (10)$$

Whence, since  $\Phi$  is subadditive and continuous,  $f$  is also continuous and  $\lim_{K \ni k \rightarrow \infty} x^k = \bar{x}$ , we get

$$\begin{aligned} \limsup_{K \ni k \rightarrow \infty} \Phi(\rho_k P(x^k)) &\leq \limsup_{K \ni k \rightarrow \infty} \left[ \Phi(f(x^k) + \rho_k P(x^k)) + \Phi(-f(x^k)) \right] \\ &\leq \hat{v} + \Phi(-f(\bar{x})). \end{aligned} \quad (11)$$

So, from the properties of  $\limsup$ , there exists  $K_1 \subset K$  and  $k_0 \in K_1$  such that

$$\Phi(\rho_k P(x^k)) < |\hat{v} + \Phi(-f(\bar{x}))| + 1 \quad \text{for all } K_1 \ni k \geq k_0.$$

Since  $\rho_k P(x^k) \geq 0$  for all  $k$  and  $\Phi$  satisfies condition (Q), from the above inequality we see that  $\|\rho_k P(x^k)\| \leq T$  for some  $T > 0$  and all  $K_1 \ni k \geq k_0$ . Using the fact that  $\{\rho_k\}$  is a divergent sequence of positive real numbers, it follows that

$$\limsup_{K_1 \ni k \rightarrow \infty} P(x^k) = 0.$$

Whence, the continuity of  $P$  yields

$$P(\bar{x}) = 0,$$

which means that  $\bar{x} \in D$ . Therefore, since  $\Phi$  is  $w$ -increasing and  $\rho_k P(x^k) \geq 0$  for all  $k$ , letting  $K \ni k \rightarrow \infty$  in (10), we obtain

$$\Phi(f(\bar{x})) \leq \hat{v}.$$

Since  $\hat{v} = \inf_{x \in D} \Phi(f(x))$  and  $\bar{x} \in D$ , we conclude that  $\bar{x} \in \operatorname{argmin}_{x \in D} \Phi(f(x))$  and, once again, the result follows from item 1 of Lemma 3.2.

The proof for the strong version of *MEPM* is formally identical to the one we just saw, but using item 2 instead of item 1 of Lemma 3.2.  $\square$

The following result establishes that both versions of *MEPM* are convergent whenever  $\Phi \circ f$  has a strict minimizer in  $D$ , which happens, for instance, if this composition is strictly convex and coercive.

**Corollary 5.2.** *Assume that  $\{x^k\} \subset \mathbb{R}^n$ , as in the first part of Theorem 5.1, has an accumulation point and that the auxiliary function  $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}$  is such that  $x \mapsto \Phi(f(x))$  has a strict minimizer  $\bar{x}$  in  $D$ . Then  $x^k \rightarrow \bar{x}$  and  $\bar{x}$  is a weak Pareto solution for problem (1). If  $\Phi$  is an  $s$ -type or an  $s$ -subadditive function, the generated sequence converges to  $\bar{x}$ , a Pareto optimal solution of (1).*

*Proof.* From Theorem 5.1 and the strict optimality of  $\bar{x} \in D$ , any subsequential limit of  $\{x^k\}$  is equal to  $\bar{x}$  and so  $x^k \rightarrow \bar{x}$ , a weak Pareto optimal solution for (1). The strong convergence result also follows from Theorem 5.1.  $\square$

## 5.1 Local implementation of *MEPM* and its convergence

We now sketch a practical implementation of both, strong and weak, versions of *MEPM*. Take  $P$ , an external penalty function for  $D$ , a  $w$ -type or  $w$ -subadditive auxiliary function  $\Phi$ , an increasingly divergent sequence of positive penalty parameters  $\{\rho_k\}$  and  $V \subset \mathbb{R}^n$ , a compact set with nonempty interior (e.g., a closed ball). Let

$$x^k \in \operatorname{argmin}_{x \in V} \Phi(f(x) + \rho_k P(x)), \quad k = 1, 2, \dots \quad (12)$$

The strong version of this local implementation of the method is formally identical to the weak one, but with a strong type or strongly subadditive auxiliary function  $\Phi$ .

Note that, by Weierstrass Theorem,  $\operatorname{argmin}_{x \in V} \Phi(f(x) + \rho_k P(x)) \neq \emptyset$ , and so  $x^k$  always exists for any  $k$ . Therefore, differently from the global *MEPM*, its local variant does not require any additional assumption on the penalized objective functions and/or on the constraint set in order to be well-defined.

Assuming the existence of an isolated local minimizer of  $\Phi \circ f$  within  $D$ , we will prove that both the weak and the strong versions of the local method are fully convergent to weak Pareto and Pareto optimal solutions, respectively.

**Theorem 5.3.** *Suppose that  $\Phi: \mathbb{R}^n \rightarrow \mathbb{R}$  is a  $w$ -type or a  $w$ -subadditive auxiliary function and that  $\bar{x} \in D \subseteq \mathbb{R}^n$  is a strict local minimizer of  $\Phi \circ f$  in  $D$ , say  $\bar{x} = \operatorname{argmin}_{x \in U \cap D} \Phi(f(x))$ , for some vicinity  $U \subset \mathbb{R}^n$  of  $\bar{x}$ . Let  $\{x^k\} \subset \mathbb{R}^n$  be a sequence generated by local *MEPM* implemented with a penalty function  $P: \mathbb{R}^n \rightarrow \mathbb{R}_+^m$ , a parameters sequence  $\{\rho_k\} \subset \mathbb{R}_{++}$ , the auxiliary function  $\Phi$  and  $V \subset U$ , a compact vicinity of  $\bar{x}$ . Then  $x^k \rightarrow \bar{x}$  and  $\bar{x}$  is a weak-Pareto optimal solution for problem (1). If  $\Phi$  is an  $s$ -type or an  $s$ -subadditive auxiliary function, then  $x^k \rightarrow \bar{x}$  and  $\bar{x}$  is a Pareto optimal solution for problem (1).*

*Proof.* The sequence  $\{x^k\} \subset V$  has an accumulation point  $\tilde{x} \in V$  because  $V$  is compact. As in item 2 of Theorem 5.1, we see that  $\tilde{x} \in \operatorname{argmin}_{x \in D \cap V} \Phi(f(x))$ . Since  $\bar{x}$  is the unique minimizer of  $\Phi \circ f$  within  $D \cap U$  and  $V \subset U$ , we conclude that  $\tilde{x} = \bar{x}$ ; so the unique accumulation point of  $\{x^k\}$  is  $\bar{x}$  and the proof is complete.  $\square$

## 5.2 Comparison with White's method

The method proposed by White [18] for solving problem (1) applies only to the case in which the constraint set  $D$  is compact. Moreover, at each iteration, it requires the computation of a Pareto solution of the penalized subproblem. For  $D$  compact,  $P = (\hat{P}, \dots, \hat{P}) \in \mathbb{R}^m$ , with  $\hat{P}$  a scalar-valued external penalty function for  $D$ , by item 2 of Lemma 3.2, the strong version of *MEPM* falls in the scheme proposed by White [18]. From a computational point of view, penalty functions as those proposed by White, i.e.,  $\hat{P}e$ , with  $e^\top = (1, \dots, 1)$ , seem to be a natural choice. However, in the case that some components  $f_i$  of the objective function are in quite different scales, we may compensate this drawback by using appropriate choices for  $P_i$ . For general problems, our proposal is another possibility of a penalty approach.

So, *MEPM* extends White's procedure, since it admits larger classes of constraint sets and of penalty functions. Moreover, in practical terms, iteration (4) has the advantage to give a precise definition of the whole sequence. In order to guarantee convergence to a Pareto optimal solution, White's procedure requires convexity of the constraint set, as well as of the objective and the penalty functions. While the convergence results for *MEPM* are based on the existence of accumulation points of the generated sequence, as well as on the existence of (local) strict minimizers of the auxiliary function, both conditions required in Zangwill's procedure for scalar optimization.

## 6 Other families of auxiliary functions

Up to now, we studied weak and strong versions of global and local *MEPM*, i.e., implemented with a  $w$ -type or a  $w$ -subadditive,  $s$ -type or  $s$ -subadditive auxiliary function  $\Phi$ . Let us now analyze why we need these kind of functions. The reason is quite simple: as we know, by Lemma 3.2, the monotonic behavior of the auxiliary function guarantees that the minimizers of  $\Phi \circ f$  within  $D$  are weak Pareto or Pareto optimal points for problem (1); while (P) jointly with the continuity of  $\Phi$  or (Q) together with the subadditivity and continuity of  $\Phi$  allow us to prove that  $\{\rho_k P(x^k)\}_{k \in K}$  has a bounded subsequence (see (9) and (11)). So, proceeding as in the proofs of Theorems 5.1 and 5.3, we can see that the accumulation points of  $\{x^k\}$  are feasible and then establish that these limits are in  $\operatorname{argmin}_{x \in D} \Phi(f(x))$ , which means that they are (weak or strong) optima for problem (1).

We just used those four kind of auxiliary functions (of weak or strong-type, weakly or strongly subadditive) due to the fact that they seem very natural and, mainly, because they allowed us to exhibit lots of very simple examples that can be used in practical implementations of the method. Moreover, we did not want to complicate the convergence results statements (and, perhaps their proofs too) by using larger classes of auxiliary functions.

In order to give an example of application of the method, let us now show other (parametrized) auxiliary functions classes that can be used in the method and such that, with them, our convergence results also hold. First, let us study a class of  $w$ -increasing auxiliary functions that are neither of  $w$ -type nor  $w$ -subadditive. Take  $\omega \in \mathbb{R}^m$ ,  $\alpha \in \mathbb{R}_{++}$ ,



a  $w$ -type function  $\Phi: \mathbb{R}^m \rightarrow \mathbb{R}$ , and define

$$\Phi_{\alpha,\omega}: \mathbb{R}^m \rightarrow \mathbb{R}, \quad \Phi_{\alpha,\omega}(u) := \Phi(\alpha u + \omega).$$

Clearly, for any  $(\alpha, \omega)^\top \in \mathbb{R}_{++} \times \mathbb{R}_+^m$ , the function  $\Phi_{\alpha,\omega}$  is continuous,  $w$ -increasing and it satisfies property (Q). But, in view of the presence of  $\alpha$ , it does not necessarily satisfy property (P). So, in general, these functions are not of  $w$ -type and, in principle, they are not subadditive. Nevertheless, even for  $\omega \in \mathbb{R}^m \setminus \mathbb{R}_+^m$ , the function  $\Phi_{\alpha,\omega}$  can be used in the method and the produced sequence will have the same convergence properties as those generated with  $w$ -type or  $w$ -subadditive auxiliary functions. Indeed, all we need to show is that  $\rho_k P(x^k) \leq M$  for all  $k \in K$ , an infinite subset of  $\{1, 2, \dots\}$ , and some  $M > 0$ , where  $x^k \in \operatorname{argmin}_{x \in \mathbb{R}^n} \Phi_{\alpha,\omega}(f(x) + \rho_k P(x))$  for all  $k = 1, 2, \dots$ . As in item 2 of Theorem 5.1, assume that  $\lim_{K \ni k \rightarrow \infty} x^k = \bar{x}$ . Since  $\Phi$  satisfies property (P), for any  $\tilde{x} \in D$ , all  $j = 1, \dots, m$  and all  $k \in K$ , we have

$$\begin{aligned} \alpha(f_j(x^k) + \rho_k P_j(x^k)) + \omega_j &\leq \Phi(\alpha[f(x^k) + \rho_k P(x^k)] + \omega) \\ &= \Phi_{\alpha,\omega}(f(x^k) + \rho_k P(x^k)) \\ &\leq \Phi_{\alpha,\omega}(f(\tilde{x})). \end{aligned}$$

Hence,

$$0 \leq \limsup_{k \in K} \rho_k P_j(x^k) \leq \frac{1}{\alpha} [\Phi_{\alpha,\omega}(f(\tilde{x})) - \omega_j] - f_j(\bar{x})$$

for all  $j = 1, \dots, m$  and any  $\tilde{x} \in D$ . Whence,  $\{\rho_k P(x^k)\}_{k \in K}$  has a bounded subsequence. The convergence and the optimality of the subsequential limit point follow as in the proof of Theorem 5.1, item 2.

Now we analyze a particular case of those new auxiliary functions families. Observe that for  $\Phi(u) := \max_{i=1, \dots, m} \{u_i\}$ , which is of  $w$ -type and  $w$ -subadditive, the function given by  $\Phi_{\alpha,\omega}(u) := \max_{i=1, \dots, m} \{\alpha u_i + \omega_i\}$  is continuous,  $w$ -increasing and satisfies property (Q), for any  $\alpha \in \mathbb{R}_{++}$  and  $\omega \in \mathbb{R}^m$ , but, due to the presence of  $\omega \in \mathbb{R}^m$ , it is not subadditive and, for  $\omega \in \mathbb{R}^m \setminus \mathbb{R}_+^m$ , in principle it is not of  $w$ -type. However, as we will now show, this kind of functions can be used in the local version of *MEPM* and the results of Theorem 5.3 are still valid. Let  $x^k \in \operatorname{argmin}_{x \in V} \Phi_{\alpha,\omega}(f(x) + \rho_k P(x))$ , where  $V \subset \mathbb{R}^n$  is a compact vicinity of a point  $\bar{x} \in \mathbb{R}^n$ . Assume that a subsequence  $\{x^k\}_{k \in K}$  is such that  $x^k \rightarrow \bar{x}$  as  $K \ni k \rightarrow \infty$ . As we know, all we need to show is that  $\{\rho_k P(x^k)\}_{k \in K}$  has a bounded subsequence. For any  $\tilde{x} \in D \cap V$ , we have

$$\begin{aligned} \limsup_{K \ni k \rightarrow \infty} \Phi_{\alpha,\omega}(\rho_k P(x^k)) &\leq \limsup_{K \ni k \rightarrow \infty} \left[ \Phi_{\alpha,\omega}(f(x^k) + \rho_k P(x^k)) + \Phi(-\alpha f(x^k)) \right] \\ &\leq \Phi_{\alpha,\omega}(f(\tilde{x})) + \Phi(-\alpha f(\tilde{x})), \end{aligned} \tag{13}$$

where we used the subadditivity of  $\Phi$ . We conclude that all results of Theorem 5.3 hold for  $\{x^k\}$  generated with  $\Phi_{\alpha,\omega}$  as auxiliary function.

Note that, in the very same way, we can see that any auxiliary function of the form  $u \mapsto \Phi_{\alpha,\omega}(u) := \Phi(\alpha u + \omega)$ , where  $\Phi$  is a  $w$ -increasing continuous function such that it

satisfies  $\Phi_{\alpha,\omega}(u+v) \leq \Phi_{\alpha,\omega}(u) + \Gamma(v)$ , for some continuous  $\Gamma: \mathbb{R}^m \rightarrow \mathbb{R}$ , will also allow us to prove that  $\{\Phi_{\alpha,\omega}(\rho_k P(x^k))\}_{k \in K}$  has a bounded subsequence, where, of course,  $x^k \in \operatorname{argmin}_{x \in V} \Phi_{\alpha,\omega}(f(x) + \rho_k P(x))$ . Actually, for  $\Phi(u) = \max_{i=1,\dots,m} \{u_i\}$ , we have  $\Gamma = \Phi$  (see (13)).

So,  $w$ -increasing continuous auxiliary functions, different from those used in (4) and in (12), can also be used, by means of iterations like (4) or (12), in order to produce sequences  $\{x^k\}$  which enjoy good convergence properties. (And, clearly, we could have shown similar examples with strong type or strongly subadditive auxiliary functions.) We just need them to be monotonic and such that, in the global case, the iterates exist and, in both global and local cases, they allow us to prove that  $\{\rho_k P(x^k)\}$  has a bounded subsequence.

Let us finish this section with an application of the method for sequences produced with auxiliary functions as those we have just examined. Actually, we will exhibit a very simple instance of problem (1), for which there exists a family of auxiliary functions  $\{\Psi_\omega\}_{\omega \in \Omega}$ ,  $\Omega \subset \mathbb{R}^m$ , such that, using local *MEPM* implemented with these functions, any penalty function and any parameter sequence, by varying  $\omega \in \Omega$ , we can retrieve the whole optimal set.

**Example 6.1.** Consider  $n = 1$ ,  $m = 2$ ,  $D = [-2, +\infty)$  and  $f: \mathbb{R} \rightarrow \mathbb{R}^2$  defined by  $f_1(x) := x^2 + 1$ ,  $f_2(x) := x^2 - 2x + 1$ . In Fig. 1a, we see that, in the interval  $[0, 1]$ , whenever  $f_2$  decreases,  $f_1$  increases and this happens only in this interval, that is to say,  $[0, 1]$  is the weak Pareto optimal set. We will apply *MEPM* with  $\Psi_\omega(u) := \Phi_{1,\omega}(u) = \max_{i=1,2} \{u_i + \omega_i\}$ ,  $\omega \in \mathbb{R}^2$ . First of all, note that, in Fig. 1a, we can also verify that  $f_1(x) \leq f_2(x)$  if and only if  $x \leq 0$ , so  $x \mapsto \Psi_0(f(x))$  has a strict minimizer at  $x = 0$ . Let us investigate  $\operatorname{argmin}_{x \in D} \Psi_\omega(f(x))$  not just for  $\omega = 0$ . It is easy to see that  $x \mapsto \Psi_\omega(f(x))$  has a unique minimizer in  $[0, 1]$  at the sole point  $\bar{x} \in [0, 1]$  where  $f_1(\bar{x}) + \omega_1 = f_2(\bar{x}) + \omega_2$ , that is to say  $\omega_1 = \omega_2 - 2\bar{x}$ . Taking  $\omega^\top := (-2\bar{x}, 0)$ , we have  $\operatorname{argmin}_{x \in D} \Phi_\omega(f(x)) = \{\bar{x}\}$ , since,  $f_1(x) + \omega_1 \leq f_2(x) + \omega_2$  if and only if  $x \leq \bar{x}$  (see Fig. 1b, 1c, 1d for  $\bar{x} = 0, 0.5, 1$ , respectively).

Following the proof of Theorem 5.3, for the auxiliary function  $\Psi_\omega$ , with  $\omega^\top := (-2\bar{x}, 0)$ , where  $\bar{x} \in [0, 1]$ , any vector external penalty function  $P$  for  $D$ , as well as any  $V \subset \mathbb{R}$  compact vicinity of  $\bar{x}$  and any parameter sequence  $\{\rho_k\}$ , the generated sequence  $\{x^k\}$  converges to  $\bar{x}$ , a weak-Pareto optimal point for  $f$  in  $D$ . This means that, by varying the parameter  $\omega \in \Omega := [-2, 0] \times \{0\}$ , the family of auxiliary functions  $\{\Psi_\omega\}_{\omega \in \Omega}$  allows us to retrieve the whole weak Pareto optimal set of the original problem.

Of course, this is an *ad hoc* example, but it may be useful in order to investigate when do we have auxiliary functions families such that by varying the parameters we can obtain the whole optimal frontier by means of the corresponding sequences.

## 7 Final remarks

For the multicriteria optimization setting, we developed an extension of Zangwill's scalar-valued method. As expected, the multiobjective convergence results are not stronger than

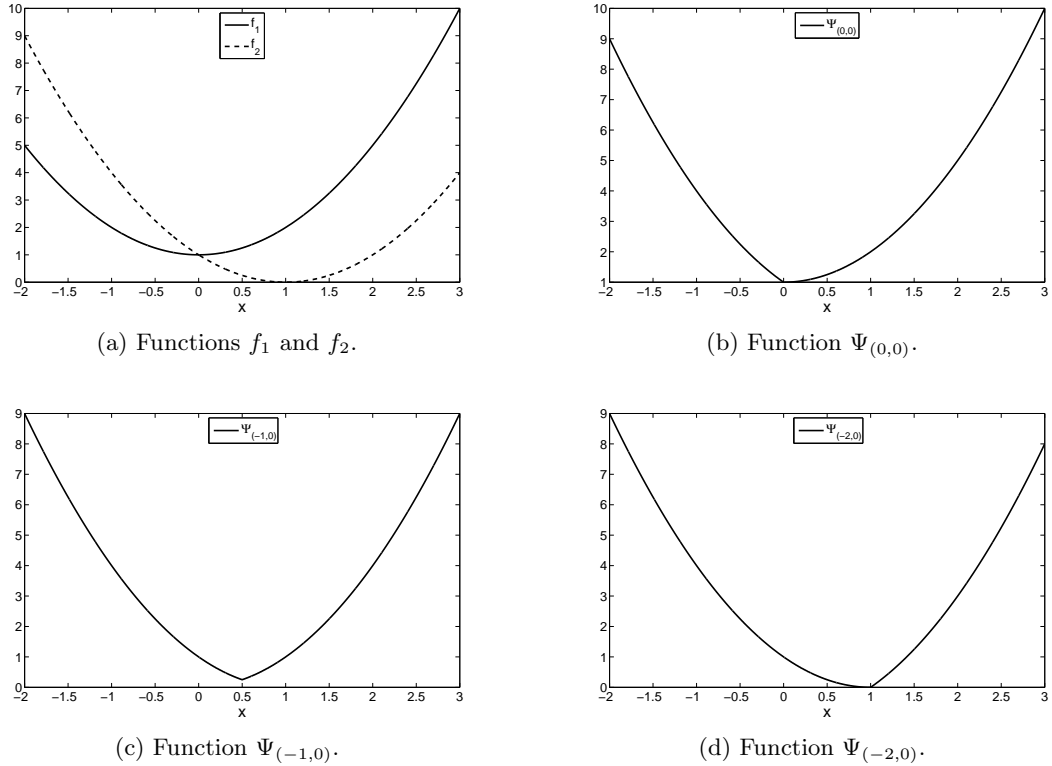


Figure 1: Objective and auxiliary functions from Example 6.1.

those for the classical method. Actually, when restricted to single objective optimization, they are both the same.

An important subject to be examined, which will be left for a future research, is when we can obtain the whole Pareto (weak Pareto) frontier by using *MEPM*. Even though we do not intend to deepen in this matter here, let us make some comments on it. Example 14 suggests that Theorem 5.3 can shed some light on the subject: the fact that *MEPM* only converges to minimizers of the scalar representations induced by the auxiliary functions — which could be considered a drawback of the method — can be very useful in order to study necessary and/or sufficient conditions for the existence of auxiliary functions families  $\{\Phi_\omega\}_{\omega \in \Omega}$  with such property.

Another matter worth to be studied is the following. Recall that the main hypothesis in Theorem 5.3 is that  $\Phi \circ f$  has a strict local minimizer within  $D$ . This convergence result may also be true under a weaker condition, namely whenever  $S := \operatorname{argmin}_{x \in U \cap D} \Phi(f(x))$  is an isolated set of  $V^* := \{x \in \mathbb{R}^n : \Phi(f(x)) = v^*\}$ , where  $v^* = \inf_{x \in D} \Phi(f(x))$ , which means that there exists a closed set  $G \subset \mathbb{R}^n$  such that  $\emptyset \neq S \subset \operatorname{int}(G)$  and  $G \setminus S \subset \mathbb{R}^n \setminus V^*$ , where  $\operatorname{int}(G)$  stands for the interior of  $G$ . Finally, it would be also interesting to study the generalization of *MEPM* to the vector optimization case.

**Acknowledgements:** We would like to thank the anonymous referees for their suggestions which improved the original version of the paper. We are also thankful to Alfredo N. Iusem and Benar F. Svaiter for valuable discussions.

## References

- [1] H. Bonnel, A. N. Iusem, and B. F. Svaiter. Proximal methods in vector optimization. *SIAM Journal on Optimization*, 15(4):953–970, 2005.
- [2] E. Carrizosa and J. B. G. Frenk. Dominating sets for convex functions with some applications. *Journal of Optimization Theory and Applications*, 96(2):281–295, 1998.
- [3] H. Eschenauer, J. Koski, and A. Osyczka. *Multicriteria Design Optimization*. Springer, Berlin, 1990.
- [4] J. Fliege, L. M. Graña Drummond, and B. F. Svaiter. Newton’s method for multiobjective optimization. *SIAM Journal on Optimization*, 20(2):602–626, 2009.
- [5] J. Fliege and B. F. Svaiter. Steepest descent methods for multicriteria optimization. *Mathematical Methods of Operations Research*, 51(3):479–494, 2000.
- [6] Y. Fu and U. M. Diwekar. An efficient sampling approach to multiobjective optimization. *Annals of Operations Research*, 132(1-4):109–134, 2004.
- [7] E. H. Fukuda and L. M. Graña Drummond. On the convergence of the projected gradient method for vector optimization. *Optimization*, 60(8-9):1009–1021, 2011.
- [8] E. H. Fukuda and L. M. Graña Drummond. Inexact projected gradient method for vector optimization. *Computational Optimization and Applications*, 54(3):473–493, 2013.
- [9] L. M. Graña Drummond and A. N. Iusem. A projected gradient method for vector optimization problems. *Computational Optimization and Applications*, 28(1):5–29, 2004.
- [10] L. M. Graña Drummond, F. M. P. Raupp, and B. F. Svaiter. A quadratically convergent Newton method for vector optimization. *Optimization*, 63(5):661–677, 2014.
- [11] L. M. Graña Drummond and B. F. Svaiter. A steepest descent method for vector optimization. *Journal of Computational and Applied Mathematics*, 175(2):395–414, 2005.
- [12] M. Gravel, J. R. Martel, W. Price, and R. Tremblay. A multicriterion view of optimal resource allocation in job-shop production. *European Journal of Operations Research*, 61:230–244, 1992.
- [13] J. Jahn. *Vector Optimization – Theory, Applications and Extensions*. Springer, Erlangen, 2003.

- [14] T. M. Leschine, H. Wallenius, and W. A. Verdini. Interactive multiobjective analysis and assimilative capacity-based ocean disposal decisions. *European Journal of Operational Research*, 56:278–289, 1992.
- [15] D. T. Luc. Theory of vector optimization. In *Lecture Notes in Economics and Mathematical Systems*, 319, Berlin, 1989. Springer.
- [16] D. G. Luenberger. *Linear and Nonlinear Programming*. Kluwer Academic Publishers, Boston, 2003.
- [17] M. Tavana. A subjective assessment of alternative mission architectures for the human exploration of mars at NASA using multicriteria decision making. *Computational and Operations Research*, 31:1147–1164, 2004.
- [18] D. J. White. Multiobjective programming and penalty functions. *Journal of Optimization Theory and Applications*, 43(4):583–599, 1984.
- [19] D. J. White. Epsilon-dominating solutions in mean-variance portfolio analysis. *European Journal on Operations Research*, 105:457–466, 1998.
- [20] W. I. Zangwill. Non-linear programming via penalty functions. *Management Science*, 13(5):344–358, 1967.