A projected gradient method for vector optimization problems

L. M. Graña Drummond* A. N. Iusem†

January 4, 2002

Abstract

We consider extensions of the projected gradient gradient method to vector optimization, which work directly with vector-valued functions, without using scalar-valued objectives. We provide a direction which adequately subtitutes for the projected gradient, and establish results which mirror those available for the scalar-valued case, namely stationarity of the cluster points (if any) without convexity assumptions, and convergence of the full sequence generated by the algorithm to a weakly efficient optimum in the convex case, under mild assumptions. We also prove that our results still hold when the search direction is only approximately computed.

Key words: Pareto optimality, weak efficiency, projected gradient methods, convexity with respect to cones, quasi-Fejér convergence.

*Programa de Engenharia de Sistemas de Computação, COPPE–UFRJ, CP 68511, Rio de Janeiro–RJ, 21945–970, Brazil (**lmgd@cos.ufrj.br**). The work of this author was supported by FAPERJ under the grant E-26/152.107/1999-Bolsa

†Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil (**iusp@impa.br**). The work of this author was partially supported by CNPq grant no. 301280/86

1 Introduction

1.1 The problem

Given a closed, convex and pointed cone K in a finite dimensional vector space V, we consider the partial order "$\preceq$" in V defined as $z \preceq z'$ (alternatively, $z' \succeq z$) if and only if $z' - z \in K$. Assuming that K has nonempty interior, we say that $z \prec z'$ (alternatively, $z' \succ z$) if and only if $z' - z \in \mathrm{int}(K)$. Consider now another finite dimensional vector space V', a closed and convex subset C of V, and a continuously differentiable function $f : V' \to V$. We are interested in the problem

$$\min\nolimits_K f(x) \tag{1}$$

$$\text{s.t. } x \in C, \tag{2}$$

with the following meaning: a vector $x^* \in V'$ is a weakly efficient solution of (1)–(2) if and only if x^* belongs to C and there exists no $x \in C$ such that $f(x) \prec f(x^*)$.

We introduce next some notation. We will refer to problem (1)–(2) as a *vector-valued* optimization problem. In the particular case of $V = \mathbb{R}^m$ and $K = \mathbb{R}^m_+$ (i.e. K is the non-negative orthant of $\mathbb{R}^m$), $\preceq$ is the usual partial order in $\mathbb{R}^m$, i.e. the component-wise order. We will refer to this case as the *multiobjective* optimization problem. When $m = 1$, $\preceq$ reduces furtherly to the usual linear order in $\mathbb{R}$, and the problem will be called a *scalar-valued* optimization problem.

We consider also an inner product $\langle \cdot, \cdot \rangle$ in V, and K^* will denote the positive polar cone of K, i.e. $K^* = \{y \in V : \langle y, z \rangle \geq 0 \, \forall x \in K\}$.

We say that f as above is K-convex if and only if

$$f(\lambda x + (1 - \lambda)x') \preceq \lambda f(x) + (1 - \lambda) f(x') \tag{3}$$

for all $x, x' \in V'$ and all $\lambda \in (0, 1)$.

We will be concerned with methods for finding a weakly efficient solution of the vector-valued optimization problem (1)–(2). We emphasize that we are not attempting to find, or somehow characterize, the set of all efficient solutions, weak or otherwise. It has been observed long ago (see e.g. [6], [9] for the multiobjective case), that the vector-valued optimization problem can be in a certain sense reduced to a scalar-valued one in the following way: take any $y \in K^*$ and define $\varphi_y : V' \to \mathbb{R}$ as $\varphi_y(x) = \langle y, f(x) \rangle$. Then, every solution x^* of min $\varphi_y(x)$ subject to $x \in C$ is a weakly efficient solution of (1)–(2). It could be thought that this result settles the computational issue of solving the vector-valued problem: any algorithm known to work for scalar-valued optimization problems could do the job, by solving the *scalarized* problem min $\varphi_y(x)$ s.t. $x \in C$. The limitation of this *scalarization* approach lies in the fact that φ_y might lack minimizers in C, even when the vector-valued problem has solutions (see e.g. [7]). It is a very difficult task to find a $y \in K^*$ such that the scalarized problem is guaranteed to have solutions, and in fact it is not easier to ascertain that such a y exists, even when f is K-convex (unless $f(C)$ is convex). Thus, it is relevant to present algorithms which deal directly with the vector-valued problem, without performing any scalarization. Extensions of the steepest descent method for unconstrained scalar-valued optimization to vector-valued problems have been introduced in [7] for the multiobjective case and in [11] for the case of a general cone K. These extensions will be discussed in subsection 1.4. In this paper we deal with the constrained vector-valued optimization problem (1)–(2), and our aim is to extend to constrained vector-valued problems the projected gradient method for constrained scalar-valued ones. We start by recalling, in the following three subsections, existing convergence results related to the steepest descent method for unconstrained scalar-valued problems, to the projected gradient method for constrained scalar-valued problems, and to the extensions of the steepest descent method for unconstrained vector-valued problems.

1.2 The steepest descent method

Given a continuously differentiable $f : \mathbb{R}^n \to \mathbb{R}$, the *steepest descent* method generates a sequence $\{x^k\} \in \mathbb{R}^n$ through

$$x^{k+1} = x^k - \beta_k \nabla f(x^k), \tag{4}$$

where β_k is some positive real number. Several choices are available for β_k. The first one is to set the β_k's exogenously, and a relevant option is

$$\beta_k = \frac{\alpha_k}{\|\nabla f(x^k)\|}, \tag{5}$$

with

$$\sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty. \tag{6}$$

Other options consist of performing an exact line minimization, i.e.

$$\beta_k = \mathrm{argmin}_{\beta>0} f(x^k - \beta \nabla f(x^k)), \tag{7}$$

or an inexact linesearch, e.g. following an Armijo rule, namely

$$\beta_k = \bar{\beta} 2^{-\ell(k)} \tag{8}$$

with

$$\ell(k) = \min\left\{ j \in \mathbb{Z}_{\geq 0} : f(x^k - \bar{\beta} 2^{-j} \nabla f(x^k)) \leq f(x^k) - \sigma \bar{\beta} 2^{-j} \left\| \nabla f(x^k) \right\|^2 \right\}, \tag{9}$$

for some $\bar{\beta} > 0$, $\sigma \in (0, 1)$.

The basic convergence result on this method (and in general on descent direction methods), under either exact linesearches or Armijo searches, derives from Zangwill's global convergence theorem (see [19]) and establishes that every cluster points $\bar{x}$ of $\{x^k\}$, if any, is stationary, i.e. such that $\nabla f(\bar{x}) = 0$. In order to ensure existence of cluster points, it is necessary to assume that the starting iterate x^0 belongs to a bounded level set of f (see [17] for this and other related results). The situation is considerably better when f is convex: it is possible to prove convergence of the whole sequence to a minimizer of f under the sole asumption of existence of minimizers (i.e. without any additional assumption on boundedness of level sets). Results of this kind for the convex case can be found in [5], [10] and [16] for the method with exogenously given β_k's satisfying (5)–(6), in [13] for the method with exact lineasearches as in (7), and in [2], [4] and [14] for the method with the Armijo rule (8)–(9). We observe that in the case of β_k's given by (5)–(6) the method is not in general a descent one, i.e. it is not guaranteed that $f(x^{k+1}) \leq f(x^k)$ for all k, so that the convergence analysis of this option is harder than in the other cases.

1.3 Projected gradient methods

We consider now the problem

$$\min f(x) \tag{10}$$

$$\text{s. t. } x \in C, \tag{11}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and $C \subset \mathbb{R}^n$ is closed and convex. Convexity of C makes it posible to use the orthogonal projection $P_C : \mathbb{R}^n \to C$ onto C for obtaining feasible directions which are also descent ones; namely a step is taken from x^k in the direction of $-\nabla f(x^k)$, the resulting vector is projected onto C, and the direction from x^k to this projection has the above mentioned properties. A formal description of the algorithm, called the *projected gradient* method, is the following:

$$z^k = P_C(x^k - \beta_k \nabla f(x^k)), \tag{12}$$

$$x^{k+1} = x^k + \gamma_k(z^k - x^k), \tag{13}$$

where β_k, γ_k are positive stepsizes, for which, again, several choices are possible. Before discussing them, we mention that in the unconstrained case, i.e. $C = \mathbb{R}^n$, then method given by (12)–(13) with $\gamma_k = 1$ for all k reduces to (4). Following [3], we will focus in three strategies for the stepsizes:

i) Armijo search along the feasible direction: $\beta_k = \widehat{\beta} > 0$ for all k and γ_k determined with an Armijo rule, namely (8)–(9) with $\bar{\beta} = 1$ and γ_k substituting for β_k.

ii) Armijo search along the boundary of C: $\gamma_k = 1$ for all k and β_k determined through (8) and the following two equations instead of (9):

$$\ell(k) = \min \left\{ j \in \mathbb{Z}_{\geq 0} : f(z^{k,j}) \leq f(x^k) - \sigma \bar{\beta} 2^{-j} \left\| \nabla f(x^k) \right\|^2 \right\},$$

with

$$z^{k,j} = P_C(x^k - \bar{\beta} 2^{-j} \nabla f(x^k)).$$

iii) Exogenous stepsize before projecting: β_k given by (5)–(6) and $\gamma_k = 1$ for all k.

Several comments are in order. First, observe that in the unconstrained case ($C = \mathbb{R}^n$) options (i) and (ii) reduce to the steepest descent method (4) with the Armijo rule given by (8)–(9), while option (iii) reduces to (4) with the β_k's given by (5)–(6). Secondly, note that option (ii) requires a projection onto C for each step of the inner loop resulting from the Armijo search, i.e. possibly many projections for each k, while option (i) demands only one projection for each outer step, i.e. for each k. Thus, option (ii) is competitive only when P_C is very easy to compute (e.g. when C is a box or a ball). Finally, we mention that option (iii), as its counterpart in the unconstrained case, fails to be a descent method.

Without assuming convexity of f, the convergence results for these methods closely mirror the ones for the steepest descent method in the unconstrained case: cluster points may fail to exist, even when (10)–(11) has solutions, but if they exist, they are stationary and feasible, i.e. $\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0$, $\bar{x} \in C$ for all cluster point $\bar{x}$ of $\{x^k\}$ and all $x \in C$. These results can be found in Section 2.3.2 of [3]; for the case of option (ii), they are based upon the results in [8].

When f is convex, the stronger results for the unconstrained case, with β_k's given by (5)–(6), have also been extended to the projected gradient method under option (iii): it has been proved in [1] that in such a case, the whole sequence $\{x^k\}$ converges to a solution of problem (10)–(11) under the sole assumption of existence of solutions. On the other hand, the situation is rather worse for options (i) and (ii): as far as we know, nothing has been proved in connection with existence and/or uniqueness of cluster points for these options assuming only convexity of f. This issue stands therefore as an interesting open problem. Of course, results of this kind are immediate under stronger hypotheses on the problem: cluster points of $\{x^k\}$ certainly exist if the intersection of C with some level set of f is nonempty and bounded, and strict convexity of f ensures uniqueness of the cluster point.

1.4 Steepest descent method for vector-valued problems

We are concerned in this subsection with the unconstrained version of problem (1)–(2). We will assume for the sake of an easier notation that $V' = \mathbb{R}^n$ and $V = \mathbb{R}^m$ with the Euclidean inner product, so that, given a continuously differentiable $f : \mathbb{R}^n \to \mathbb{R}^m$, we look for some $\bar{x} \in \mathbb{R}^n$ such that there exists no $x \in \mathbb{R}^n$ satisfying $f(x) \prec f(\bar{x})$, where "$\prec$" is induced by a closed, convex and pointed cone $K \subset \mathbb{R}^m$ with nonempty interior. The following vector-valued generalization of the steepest descent method, which avoids pre-established scalarizations, was proposed first in [7] for the multiobjective case (i.e. $K = \mathbb{R}^m_+$), and later on extended to the case of a general cone $K \subset \mathbb{R}^m$ in [11]. The algorithm makes use of a compact set G of normalized generators of the polar cone K^*, i.e. $G \subset K^*$ is compact and such that the cone generated by its convex hull is precisely K^*. Such a G always exists; one can take e.g. $G = \{y \in K^* : \|y\| = 1\}$, but in general it is possible to obtain much smaller sets: in the multiobjective case G can be taken as the canonical basis of $\mathbb{R}^m$, and, furthermore, when K is polyhedral, K^* is also polyhedral, so G can be chosen as the finite collection of the polar cone generators. From now on, for all $x \in \mathbb{R}^n$, $J_f(x)$ will denote the Jacobian matrix of f at x. The iterative step of the method is given by:

$$w^k = \operatorname{argmin}_{w \in \mathbb{R}^n} \left\{ \frac{\|w\|^2}{2} + \max_{y \in G} \left(y^t J_f(x^k) w \right) \right\}, \tag{14}$$

$$x^{k+1} = x^k + \beta_k w^k, \tag{15}$$

where β_k is found through an Armijo search, but now working with the partial order in $\mathbb{R}^m$, i.e. instead of (8)–(9) we have:

$$\beta_k = \bar{\beta} 2^{-\ell(k)} \tag{16}$$

with

$$\ell(k) = \min \left\{ j \in \mathbb{Z}_{\geq 0} : f(x^k + \bar{\beta} 2^{-j} w^k) \preceq f(x^k) + \sigma \bar{\beta} 2^{-j} J_f(x^k) w^k \right\}, \tag{17}$$

with $\bar{\beta} > 0$ and $\sigma \in (0, 1)$.

We remark that in the scalar-valued case (14)–(17) reduces to the steepest descent method given by (4) under the Armijo rule (8)–(9): in the scalar case $J_f(x^k) = \nabla f(x^k)^t$ and we can take $G = \{1\}$, so that (14) becomes

$$w^k = \operatorname{argmin}_{w \in \mathbb{R}^n} \left\{ \frac{\|w\|^2}{2} + \nabla f(x^k)^t w \right\}, \tag{18}$$

and it is immediate from (18) that $w^k = -\nabla f(x^k)$. The results in [11] attain a complete extension of those known for the steepest descent method with an Armijo rule in the scalar-valued case: assuming only continuous differentiability of f, it is proved that the algorithm is well defined, in the sense that the Armijo search is always successful (i.e. a β_k satisfying (16) and (17) exists), and that every cluster point $\bar{x}$ of $\{x^k\}$, if any, is a stationary point for problem (1) (i.e. that $\operatorname{int}(K) \cap \operatorname{Im}[J_f(\bar{x})] = \emptyset$). When f is K-convex, the corresponding stronger result is also proved in [11]: the whole sequence $\{x^k\}$ converges to a weak unconstrained minimizer x^* of f (i.e. no

$x \in \mathbb{R}^n$ exists such that $f(x) \prec f(x^*)$), under the following assumption: if $\{z^k\} \subset \mathbb{R}^n$ is such that $f(z^{k+1}) \preceq f(z^k)$ for all k then there exists $\bar{z} \in \mathbb{R}^n$ such that $f(\bar{z}) \preceq f(z^k)$ for all k. In the scalar-valued case, this condition is equivalent to requesting existence of unconstrained minimizers. Examples are given in [11] for which this assumption holds but scalarized problems, as defined above, might fail to have solutions. It is also shown in [7], [11] that the convergence results still hold when w^k in (15) is just an inexact solution of the minimization problem in (14). We will further develop this issue in Section 4.

1.5 Projected gradient methods for vector-valued minimization

We describe in this subsection the method which is introduced in this paper, devised for the constrained problem (1)–(2). It generalizes (14)–(17) to the constrained case, and (12)–(13) to the vector-valued case. The iterative step is of the form:

$$v^k = \mathrm{argmin}_{v \in C_k} \left\{ \frac{\|v\|^2}{2} + \beta_k \mathrm{max}_{y \in G} \left(y^t J_f(x^k) v \right) \right\}, \tag{19}$$

$$x^{k+1} = x^k + \gamma_k v^k, \tag{20}$$

with $C_k = C - x^k = \{u - x^k : u \in C\}$, $\{\beta_k\} \subset \mathbb{R}_{++}$ and $\{\gamma_k\} \subset (0,1]$.

Before discussing the options for the β_k's and γ_k's, we look at the specific form of (19)–(20) in the two above mentioned particular cases. Consider first the unconstrained case, i.e. $C = \mathbb{R}^n$, which implies $C_k = \mathbb{R}^n$ for all k, and look at (19) with $\beta_k = 1$ for all k. It is immediate that (19)–(20) reduce precisely to (14)–(15) with v^k, γ_k substituting for w^k, β_k. Let us consider now (19)–(20) in the constrained scalar-valued case. We take $G = \{1\}$, $J_f = (\nabla f)^t$, and (19) becomes

$$v^k = \mathrm{argmin}_{v \in C_k} \left\{ \frac{\|v\|^2}{2} + \beta_k \nabla f(x^k)^t v \right\} =$$

$$\mathrm{argmin}_{v \in C_k} \left\{ \frac{\|v\|^2}{2} + \beta_k \nabla f(x^k)^t v + \frac{\beta_k^2}{2} \left\| \nabla f(x^k) \right\|^2 \right\} =$$

$$\mathrm{argmin}_{v \in C_k} \left\{ \frac{\left\| v + \beta_k \nabla f(x^k) \right\|^2}{2} \right\} = P_{C_k} \left(-\beta_k \nabla f(x^k) \right). \tag{21}$$

Defining now $z^k = v^k + x^k$, (21) becomes

$$z^k = P_C \left(x^k - \beta_k \nabla f(x^k) \right) \tag{22}$$

and (20) becomes

$$x^{k+1} = x^k + \gamma_k(z^k - x^k). \tag{23}$$

Comparing now (22)–(23) with (12)–(13), we observe that in the scalar-valued case our method (19)–(20) coincides with the general scheme of the projected gradient method as presented in subsection 1.3.

Next we comment on the choices of β_k's and γ_k's for which our results hold. In the nonconvex case, i.e. assuming only continuous differentiability of f, we take $\beta_k = \widehat{\beta} > 0$ for all k, and then determine γ_k with the Armijo rule given by (16)–(17), taking $\bar{\beta} = 1$ and γ_k instead of β_k. This procedure coincides with scheme (i) of subsection 1.3 in the constrained scalar-valued case, and with the method in [11], i.e. (14)–(17), in the unconstrained vector-valued case, excepting for the rather immaterial presence of $\widehat{\beta}$ in the definition of v^k. Our convergence result, established in Section 2, corresponds to those in [3] and [11]: every cluster point $\bar{x}$ of $\{x^k\}$, if any, is stationary, meaning that $-\mathrm{int}(K) \cap \{J_f(\bar{x})(z - \bar{x}) : z \in C\} = \emptyset$.

For the convex case, we manage to establish convergence of the whole sequence $\{x^k\}$ to a weakly efficient solution of problem (1)–(2), under a scheme in which both the β_k's and the γ_k's change with k: we take exogenously given β_k's, of the form $\beta_k = \alpha_k/\eta_k$, where $\eta_k = \max_{y \in G} \left\| J_f(x^k)^t y \right\|$ and the α_k's satisfy (6). The γ_k's are determined through an Armijo search as in the nonconvex case, i.e. (16)–(17) with $\bar{\beta} = 1$ and γ_k instead of β_k. As discussed above, in the unconstrained vector-valued case a similar result was attained with $\beta_k = 1$ for all k and the Armijo search for the γ_k's, while in the constrained scalar-valued case the same result was proved with α_k's like in this scheme, but with $\gamma_k = 1$ for all k. The convergence analysis is presented in Section 3, where we also discuss why we need this sort of "double search", with varying β_k's and γ_k's, for achieving the result.

2 The exact method: nonconvex case

2.1 The algorithm

In this section we consider Problem (1)–(2), with $V' = \mathbb{R}^n$ and $V = \mathbb{R}^m$, and we study our algorithm defined by (19)–(20), with constant $\beta_k = \hat{\beta} > 0$ for all k, and γ_k determined by an Armijo search as in (17) with $\bar{\beta} = 1$, i.e., we take $x^0 \in C$ and the iterative step is given by:

$$v^k = \mathrm{argmin}_{v \in C_k} \left\{ \frac{\|v\|^2}{2} + \hat{\beta} \max_{y \in G} \left(y^t J_f(x^k) v \right) \right\}, \tag{24}$$

$$\ell(k) = \min \left\{ j \in \mathbb{Z}_{\geq 0} : f(x^k + 2^{-j} v^k) \preceq f(x^k) + \sigma 2^{-j} J_f(x^k) v^k \right\}, \tag{25}$$

$$x^{k+1} = x^k + \gamma_k v^k, \tag{26}$$

with $C_k = C - x^k = \{u - x^k : u \in C\}$, $\hat{\beta} > 0$, $\gamma_k = 2^{-\ell(k)}$, and $\sigma \in (0,1)$. If x^k is stationary, then the algorithm stops at iteration k.

2.2 Preliminaries

Some of the technical results in this subsection are formally similar to their unconstrained counterparts, presented in [11]. We include them here for the sake of completeness.

First of all, it is elementary that a necessary, but, in general, nonsufficient condition for weak optimality of a point $x \in C$ is stationarity, that is to say

$$-\text{int}(K) \cap J_f(x)(C - x) = \emptyset \tag{27}$$

where $J_f(x)(C - x) = \{J_f(x)(z - x) : z \in C\}$. Observe that (27) generalizes to constrained vector optimization the classical stationarity condition that holds for scalar-valued constrained optimization; indeed, if $m = 1$, $J_f(x) = \nabla f(x)^t$, and in this case (27) means precisely $\nabla f(x)^t(z - x) \geq 0 \ \forall\, z \in C$. Furthermore, it also generalizes the stationarity condition for the unconstrained vector optimization case, proposed in [11] with the label of K-criticallity, namely that condition (27) holds with $C = \mathbb{R}^n$, i.e. $-\text{int}(K) \cap \text{Im}[J_f(x)] = \emptyset$.

We point out that our methods will seek stationary points and that, as in the single-valued case, when the problem is convex, these points are exactly all the optima.

For a given nonstationary point $x \in C$, by virtue of (27), we will be able to take steps from it all along a nondegenerate interval in a certain direction, in such a way that we will obtain a strict decrease on the objective value. Moreover, we can guarantee a definite decrease, as we show in our next proposition.

Proposition 1. *Let $\sigma \in (0,1)$, $x \in C$ and $v \in C - x$ such that $J_f(x)v \prec 0$. Then there exists $\bar{\gamma} < 1$ such that $f(x + \gamma v) \prec f(x) + \sigma\gamma J_f(x)v$ for all $\gamma \in (0, \bar{\gamma}\,]$.*

Proof. Since f is differentiable, we have

$$f(x + \gamma v) = f(x) + \gamma(J_f(x)v + \varepsilon(\gamma)) \ \ \forall\, \gamma \in (0,1)\,, \tag{28}$$

where $\lim_{\gamma \to 0} \varepsilon(\gamma) = 0$.

We know that $(1 - \sigma)J_f(x)v \in -\text{int}(K)$. On the other hand, there exists $\bar{\gamma} < 1$ such that, for all $\gamma \in (0, \bar{\gamma}\,]$, $\|\varepsilon(t)\|$ is small enough so that $\varepsilon(t) + (1 - \sigma)J_f(x)v \in -\text{int}(K)$, i.e.

$$\varepsilon(t) \prec -(1 - \sigma)J_f(x)v\,, \tag{29}$$

for all $\gamma \in (0, \bar{\gamma}\,]$. Combining (28) with (29) we complete the proof.

□

The result we just proved will allow us to implement the Armijo rule on our algorithm, using a back-tracking procedure.

We need in the sequel some elementary properties of the function $x \mapsto \max_{y \in G}(y^t x)$, where $G \subset K^* \cap \{y \in \mathbb{R}^m : \|y\| = 1\}$ is compact and such that the cone spanned by its covex hull is K^*. From now on, let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be the function defined by

$$\varphi(x) := \max_{y \in G}(y^t x). \tag{30}$$

Since G spans all the directions in K^* and $K^{**} = K$ (see Theorem 14.1 in [18]), we have the following characterizations of $-K$ and $-\text{int}(K)$:

$$-K = \{y \in \mathbb{R}^m : \varphi(y) \leq 0\}, \tag{31}$$

and

$$-\text{int}(K) = \{y \in \mathbb{R}^m : \varphi(y) < 0\}. \tag{32}$$

Hence a point $x \in C$ is nonstationary (i.e. it does not satisfy condition (27)), if and only if there exists $v \in C - x$ such that

$$\varphi\left(J_f(x)v\right) < 0. \tag{33}$$

Next we present some basic properties of φ.

Proposition 2. *i) The function φ is positively homogeneous of degree 1.*

ii) $\varphi(x + z) \leq \varphi(x) + \varphi(z)$ and $\varphi(x) - \varphi(z) \leq \varphi(x - z)$ for all $x, z \in \mathbb{R}^m$.

iii) Given $x, z \in \mathbb{R}^m$, if $x \prec z$ ($x \preceq z$), then $\varphi(x) < \varphi(z)$ ($\varphi(x) \leq \varphi(z)$).

iv) The function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant $L = 1$.

Proof. (i) and (ii) are immediate.
iii) The fact that $x - z \in -\text{int}(K)$ ($x - z \in -K$) is equivalent to $\varphi(x - z) < 0$ ($\varphi(x - z) \leq 0$), by (32) (respectively, (31)). Hence, the result follows from (ii).
iv) By virtue of (ii), $\varphi(x) - \varphi(z) \leq \varphi(x - z)$ and $\varphi(z) - \varphi(x) \leq \varphi(z - x)$. Hence $|\varphi(x) - \varphi(z)| \leq \max\{\varphi(x - z), \varphi(z - x)\}$. Therefore, from (30) and the Cauchy-Schwartz inequality

$$|\varphi(x) - \varphi(z)| \leq \|x - z\|,$$

because $G \subset \{y \in \mathbb{R}^m : \|y\| = 1\}$, and the result follows. □

We study next the search direction $v(x)$ for f at $x \in C$, defined as

$$v(x) := \operatorname{argmin}_{v \in C-x} \left\{ \varphi\left(J_f(x)v\right) + \frac{\hat{\beta}}{2} \|v\|^2 \right\} \tag{34}$$

Note that the application $v \mapsto \varphi\left(J_f(x)v\right)$ is proper, closed and strongly convex. Therefore, problem (34) has a unique minimizer, which is $v(x)$. From now on, we denote the optimal value of (34) as $\theta(x)$, i.e. $\theta(x) := \varphi\left(J_f(x)v(x)\right) + \frac{\hat{\beta}}{2}\|v(x)\|^2$. Observe that, since $0 \in C - x$, we have $\theta(x) \leq 0$. Note too that, when $C = \mathbb{R}^n$, and $\hat{\beta} = 1$, we retrieve the so called K-steepest descent direction, proposed on [11]. Our next result, which characterizes stationary points in terms of $\theta(\cdot)$ and $v(\cdot)$, is essential for our convergence analysis.

Proposition 3. *For $x \in C$, the following conditions are equivalent:*

i) The point x is not stationary.

ii) $\theta(x) < 0$.

iii) $v(x) \neq 0$.

In particular, $x \in C$ is stationary if and only if $\theta(x) = 0$.

Proof. i) $\Longrightarrow$ ii) Suppose that x is nonstationary. Then, by virtue of (33), there exists $\bar{v} \in C - x$ such that $\varphi\left(J_f(x)\bar{v}\right) < 0$. Therefore, from the optimality of $v(x)$,

$$\theta(x) \leq \varphi\left(J_f(x)\tau\bar{v}\right) + \frac{\hat{\beta}}{2}\|\tau\bar{v}\|^2 = \tau\left(\varphi\left(J_f(x)\bar{v}\right) + \frac{\tau\hat{\beta}}{2}\|\bar{v}\|^2\right) \tag{35}$$

for all $\tau > 0$, where the equality follows from Proposition 2(i). If we take now in (35) τ such that $0 < \tau < -2\varphi\left(J_f(x)\bar{v}\right)/(\hat{\beta}\|\bar{v}\|^2)$, then we conclude that $\theta(x) < 0$.
ii) $\Longrightarrow$ iii) Since $\theta(x) \leq 0$, it is enough to notice that $v(x) = 0$ implies $\theta(x) = 0$.
iii) $\Longrightarrow$ i) If $v(x) \neq 0$, then $\varphi\left(J_f(x)v(x)\right) < \theta(x) \leq 0$, and the result follows from (33). □

Finally, we establish that the function θ of optimal values of (34) is continuous.

Proposition 4. *The function θ is continuous in C.*

Proof. Take $x \in C$ and let $\{x^{(k)}\} \subset C$ be a sequence such that $\lim_{k\to\infty} x^{(k)} = x$. Due to the optimality of $v(x^{(k)})$, we have, for all k,

$$\theta(x^{(k)}) \leq \varphi\left(J_f(x^{(k)})v(x)\right) + \frac{\hat{\beta}}{2}\|v(x)\|^2 .$$

Since f is continuously differentiable and, as proved in Proposition 2(iv), φ is continuous, taking $\limsup_{k\to\infty}$ on both sides of the above inequality yields

$$\limsup_{k\to\infty} \theta(x^{(k)}) \leq \theta(x)\,. \tag{36}$$

On the other hand, we have that

$$\theta(x) \leq \varphi\left(J_f(x)v(x^{(k)})\right) + \frac{\hat{\beta}}{2}\left\|v(x^{(k)})\right\|^2,$$

so

$$\begin{aligned}\theta(x) &\leq \liminf_{k\to\infty} \varphi\left(J_f(x)v(x^{(k)})\right) + \tfrac{\hat{\beta}}{2}\left\|v(x^{(k)})\right\|^2 \\ &= \liminf_{k\to\infty} \theta(x^{(k)}) + \varphi\left(J_f(x)v(x^{(k)})\right) - \varphi\left(J_f(x^{(k)})v(x^{(k)})\right) \\ &\leq \liminf_{k\to\infty} \theta(x^{(k)}) + \left\|J_f(x) - J_f(x^{(k)})\right\| \left\|v(x^{(k)})\right\|,\end{aligned} \tag{37}$$

where the second inequality holds in view of Proposition 2(iv).

Observe now that from the fact that $\theta(x) \leq 0$ for all $x \in C$ it easily follows that

$$\|v(x)\| \leq (2/\hat{\beta})\|J_f(x)\| \;\; \forall x. \tag{38}$$

So, since f is continuously differentiable and $\{x^{(k)}\}$ is convergent, we conclude from (38) that $\{v(x^{(k)})\}$ is bounded. Then, since $x^{(k)} \to x$, letting $k \to \infty$ in (37), and using once again that f is continuously differentiable, we get

$$\theta(x) \leq \liminf_{k\to\infty} \theta(x^{(k)})\,. \tag{39}$$

We complete the proof combining (36) and (39). □

2.3 Convergence of the method

First we establish feasibility of $\{x^k\}$.

Proposition 5. *If $\{x^k\}$ is a sequence defined by (24)–(26), then $\{x^k\} \subset C$.*

Proof. By induction. The initial iterate x^0 belongs to C by the initialization of the method. Assuming that $x^k \in C$, since $\gamma^k \in (0,1]$ by (25) and $v^k \in C_k = C - x^k$ by (24), we conclude that x^{k+1} belongs to C by (26) and convexity of C. □

Next we observe that if x^k is nonstationary, then $J_f(x^k)v^k \prec 0$, in which case the steplength γ_k is well defined, because $\ell(k)$ is also well defined by (25), in view of Proposition 1. Note that the sequence of objective values $\{f(x^k)\}$ decreases in the partial order induced by K, i.e. $f(x^{k+1}) \preceq f(x^k)$ for all k.

Observe also that the method can terminate with a stationary point (in a finite number of iterations) or, otherwise, the method generates an infinite sequence $\{x^k\}$. From now on we will assume that the algorithm produces an infinite sequence $\{x^k\}$ of nonstationary points. From Proposition 3, we have $\theta(x^k) \neq 0$ for all k and, in order to see that every accumulation point x^* of $\{x^k\}$ is stationary, once again in view Proposition 3, it suffices to prove that $\theta(x^*) = 0$. This will be done in our convergence theorem, which is an application of the standard convergence argument for the classical steepest descent method.

Theorem 1. *Every accumulation point, if any, of the sequence $\{x^k\}$ generated by the projected gradient method defined by* (24)-(26) *is a feasible stationary point.*

Proof. Assume that $x^* \in C$ is an accumulation point of the sequence $\{x^k\}$. Feasibility of x^* follows from Proposition 5. Let $v(x^*)$ and $\theta(x^*)$ be the solution and the optimal value of (34) at $x = x^*$, respectively, i.e. $v(x^*) = \operatorname{argmin}_{v \in C - x^*} \left\{ \varphi\left(J_f(x^*)v\right) + \frac{\hat{\beta}}{2}\|v\|^2 \right\}$ and $\theta(x^*) = \varphi\left(J_f(x^*)v(x^*)\right) + \frac{\hat{\beta}}{2}\|v(x^*)\|^2$.

According to Proposition 3, it is enough to see that $\theta(x^*) = 0$. By (25) and Proposition 2(iii) we have

$$\varphi(f(x^{k+1})) \leq \varphi(f(x^k))$$

and, since there exists a subsequence of $\{\varphi(f(x^k))\}$ converging to $\varphi(f(x^*))$, we have that

$$\lim_{k\to\infty} \varphi(f(x^k)) = \varphi(f(x^*)) \,. \tag{40}$$

On the other hand, combining (25) with Proposition 2(ii), and using Proposition 2(i), we get

$$-\sigma\gamma_k\varphi\left(J_f(x^k)v^k\right) \leq \varphi(f(x^k)) - \varphi(f(x^{k+1})) \,. \tag{41}$$

Now, using (40) and (41), as well as the fact that $J_f(x^k)v^k \prec 0$, we obtain

$$0 \leq -\lim_{k\to\infty} \sigma\gamma_k\varphi\left(J_f(x^k)v^k\right) \leq \lim_{k\to\infty} \varphi(f(x^k)) - \varphi(f(x^{k+1})) = 0 \,. \tag{42}$$

Therefore

$$\lim_{k\to\infty} \gamma_k\varphi\left(J_f(x^k)v^k\right) = 0 \,. \tag{43}$$

Consider now a subsequence $\{x^{k_j}\}_j$ converging to x^*. Taking into consideration that $\gamma_k \in (0,1]$ for all k, we have the following two possibilities:

$$\limsup_{j\to\infty} \gamma_{k_j} > 0 \,, \tag{44}$$

or

$$\limsup_{j\to\infty} \gamma_{k_j} = 0\,. \tag{45}$$

First, assume that (44) holds. Then, there exists a subsequence $\{\gamma_{k_{j_i}}\}_i$ converging to some $\gamma^* > 0$, and such that $\lim_{i\to\infty} x^{k_{j_i}} = x^*$. So, it follows from (43) that

$$\lim_{i\to\infty} \varphi\left(J_f(x^{k_{j_i}})v^{k_{j_i}}\right) = 0\,.$$

Hence, we have

$$0 = \lim_{i\to\infty} \varphi\left(J_f(x^{k_{j_i}})v^{k_{j_i}}\right) \leq \limsup_{i\to\infty} \theta(x^{k_{j_i}}) = \theta(x^*)\,,$$

where the rightmost equality follows from Proposition 4. Then, since $\theta(x^*) \leq 0$, we obtain $\theta(x^*) = 0$, and using Proposition 3, we conclude that x^* is stationary.

Now assume that (45) holds. We know that $\{x^{k_j}\}_j$ is a bounded sequence of nonstationary points, so, since v^k is bounded by $(2/\beta)\left\|J_f(x^k)\right\|$ (see (38)), the sequence of directions $\{v^{k_j}\}_j$ is bounded too. Therefore we can take subsequences $\{x^{k_{j_i}}\}_i$, $\{v^{k_{j_i}}\}_i$ and $\{\gamma_{k_{j_i}}\}_i$ converging to x^*, v^* and 0, respectively. Observe that,

$$\varphi\left(J_f(x^{k_{j_i}})v^{k_{j_i}}\right) \leq \theta(x^{k_{j_i}}) < 0\,, \tag{46}$$

where the strict inequality follows from Proposition 3 and the fact that $\{x^k\}$ is a sequence of nonstationary points. Then, taking $\lim_{i\to\infty}$ on (46), we get

$$\varphi(J_f(x^*)v^*) \leq \theta(x^*) \leq 0\,. \tag{47}$$

Take now some fixed but arbitrary positive integer q. Since $\lim_{i\to\infty} \gamma_{k_{j_i}} = 0$, for i large enough we have

$$\gamma_{k_{j_i}} < 2^{-q}\,,$$

which means that the Armijo condition does not hold at $x^{k_{j_i}}$ for $\gamma = 2^{-q}$, i.e.

$$f\left(x^{k_{j_i}} + 2^{-q}v^{k_{j_i}}\right) \npreceq f(x^{k_{j_i}}) + \sigma 2^{-q} J_f(x^{k_{j_i}})v^{k_{j_i}}\,.$$

Whence, for all positive integer i there exists $y^i \in G$ such that

$$\left\langle f\left(x^{k_{j_i}} + 2^{-q}v^{k_{j_i}}\right) - f(x^{k_{j_i}}) - \sigma 2^{-q} J_f(x^{k_{j_i}})v^{k_{j_i}}, y^i \right\rangle > 0\,. \tag{48}$$

Clearly, we do not lose generality if we assume that $\{y^i\}_i$ converges to some $\bar{y} \in G$, since G is compact. So taking limits in (48) with $i \to \infty$, we get

$$\langle f(x^* + 2^{-q}v^*) - f(x^*) - \sigma 2^{-q} J_f(x^*)v^*, \bar{y}\rangle \geq 0\,.$$

Hence,

$$\varphi\left(f(x^* + 2^{-q}v^*) - f(x^*) - \sigma 2^{-q} J_f(x^*)v^*\right) \geq 0. \tag{49}$$

Since (49) holds for any positive integer q, using Proposition 1 and (32), we conclude that $J_f(x^*)v^* \not\prec 0$, i.e.

$$0 \leq \varphi\left(J_f(x^*)v^*\right) . \tag{50}$$

Combining (47) with (50), we get that $\theta(x^*) = 0$ and, once again in view of Proposition 3, we conclude that x^* is a stationary point. □

3 The exact method: convex case

In this section we consider the projected gradient method under the assumption that f is K-convex (3), and, as in the scalar-valued case, we attempt to establish convergence of the generated sequence to a weakly efficient point under minimal assumptions. For this purpose, we consider the variant of the method with exogenously given square summable steplengths, before the computation of the feasible direction v^k (i.e. before projecting, in the scalar-valued case). For reasons to be commented upon later on, we also need the Armijo search after the computation of v^k, in order to guarantee nonincresing functional values. Thus, the algorithm under consideration replaces (24) by

$$v^k = \operatorname{argmin}_{v \in C_k} \left\{ \frac{\|v\|^2}{2} + \frac{\alpha_k}{\eta_k} \max_{y \in G}\left(y^t J_f(x^k)v\right) \right\}, \tag{51}$$

with $\{\alpha_k\} \subset \mathbb{R}^{++}$ satisfying $\sum_{k=0}^{\infty} = \infty$, $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$, and η_k given by

$$\eta_k = \max_{y \in G} \left\| J_f(x^k)^t y \right\|, \tag{52}$$

with the proviso that the algorithm stops at step k if $\eta_k = 0$. The following steps, given by (25) and (26) are kept as before, excepting that in this case we allow $\sigma \in [0, 1)$ instead of $\sigma \in (0, 1)$, i.e. we are satisfied that the Armijo search ensures just $f(x^{k+1}) \preceq f(x^k)$, because the α_k's will take care of the optimality of the cluster points.

It is easy to prove that, in the K-convex case, stationary points are weakly efficient. Also, note the new stopping criterion given by $\eta_k = 0$ implies stationarity, and henceforth weak efficiency of x^k: otherwise $-J_f(x^k)v \in \operatorname{int}(K)$ for some $v \in C_k$; taking $y^k = -J_f(x^k)v$ one has $y^k \neq 0$ because K is pointed, so that $0 < \left\|y^k\right\|^2 = -(y^k)^t J_f(x^k)v$, and therefore $J_f(x^k)^t y^k \neq 0$, in which case (52) produces $\eta_k > 0$. Finally, as in the nonconvex case, the sequence $\{x^k\}$ is feasible, with the same argument as in the proof of Proposition 5.

Define

$$\psi_k(v) = \frac{\alpha_k}{\eta_k}\varphi\left(J_f(x^k)v\right) + \frac{1}{2}\|v\|^2. \tag{53}$$

We start the convergence analysis with the following result.

Proposition 6. *Let $\{x^k\}$ and $\{v^k\}$ be as prescribed by the algorithm. Then, for all k, $\|v^k\| \le \alpha_k$ and $\|x^{k+1} - x^k\| \le \alpha_k$.*

Proof. The first order optimality conditions for $\min_{v \in C_k} \psi_k(v)$, which, in view of the convexity of ψ_k are necessary and sufficient, imply the existence of $u^k \in \partial\psi_k(v^k)$ such that

$$\langle u^k, v - v^k\rangle \ge 0 \;\forall v \in C_k\,. \tag{54}$$

Using the formula for the subdifferential of a maximum of convex functions (see [12], Volume I, Theorem VI.4.4.2), and (51), there exist $y_i^k \in G$ and $\lambda_i^k > 0$ $(1 \le i \le q(k))$ such that

$$\sum_{i=1}^{q(k)} \lambda_i^k = 1\,, \tag{55}$$

$$(y_i^k)^t J_f(x^k)v^k = \varphi\left(J_f(x^k)v^k\right) = \max_{y\in G}\left\{y^t J_f(x^k)v^k\right\} \quad (1 \le i \le q(k)), \tag{56}$$

$$u^k = v^k + \frac{\alpha_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k J_f(x^k)^t y_i^k\,. \tag{57}$$

Since $\{x^k\} \subset C$, we have that $0 \in C_k$; so, taking $v = 0$ in (54), we obtain, in view of (55)–(57),

$$\|v^k\|^2 \le -\frac{\alpha_k}{\eta_k}\sum_{i=1}^{q(k)} \lambda_i^k (y_i^k)^t J_f(x^k)v^k = -\frac{\alpha_k}{\eta_k}\max_{y\in G}\left\{y^t J_f(x^k)v^k\right\}. \tag{58}$$

Using (51) and the definition of ψ_k, we have that $\psi_k(v^k) \le \psi_k(0) = 0$, so that

$$\frac{\alpha_k}{\eta_k}\max_{y\in G}\left\{y^t J_f(x^k)v^k\right\} \le -\frac{1}{2}\|v^k\|^2 \le 0. \tag{59}$$

Combining (58) and (59), we get

$$\|v^k\|^2 \le \frac{\alpha_k}{\eta_k}\left|\max_{y\in G}\left\{y^t J_f(x^k)v^k\right\}\right| = \frac{\alpha_k}{\eta_k}\left|(\bar{y}^k)^t J_f(x^k)v^k\right|, \tag{60}$$

where $\bar{y}^k \in G$ realizes the maximum in the middle expression of (60). Now using (52) and (60), we obtain

$$\|v^k\|^2 \leq \frac{\alpha_k}{\eta_k}\|J_f(x^k)^t\bar{y}^k\|\|v^k\| \leq \alpha_k\|v^k\|\,,$$

which implies that

$$\|v^k\| \leq \alpha_k\,. \tag{61}$$

Finally, from (61), (25) and (26), we get

$$\|x^{k+1} - x^k\| = \gamma_k\|v^k\| \leq \gamma_k\alpha_k \leq \alpha_k\,,$$

and the proof is complete. □

The next proposition establishes the basic inequality satisfied by the iterates of the method.

Proposition 7. *If f is K-convex, then for all $x \in C$,*

$$3\alpha_k^2 + \|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geq \frac{2\gamma_k\alpha_k}{\eta_k}\min_{y\in G}\{[f(x^k) - f(x)]^t y\}. \tag{62}$$

Proof. Using Proposition 6 and (26), we have that

$$\alpha_k^2 + \|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geq \|x^{k+1} - x^k\|^2 + \|x^k - x\|^2 - \|x^{k+1} - x\|^2 =$$

$$2\langle x^k - x^{k+1}, x - x^k\rangle = 2\gamma_k\langle v^k, x - x^k\rangle \geq$$

$$\frac{2\gamma_k\alpha_k}{\eta_k}\sum_{i=1}^{q(k)}\lambda_i^k(y_i^k)^t J_f(x^k)(x^k - x) + 2\gamma_k\left\langle v^k, v^k + \frac{2\gamma_k\alpha_k}{\eta_k}\sum_{i=1}^{q(k)}\lambda_i^k J_f(x^k)^t y_i^k\right\rangle, \tag{63}$$

where the second inequality holds in view of (57) and (54), with $v = x - x^k \in C_k$.

On the other hand, by the gradient-like inequality for K-convex functions (see [15], Lemma 5.2), we have

$$J_f(x^k)(x^k - x) \succeq f(x^k) - f(x)\,. \tag{64}$$

Since $y_i^k \in G \subset K^*$, we get from (64)

$$(y_i^k)^t J_f(x^k)(x^k - x) \geq (y_i^k)^t[f(x^k) - f(x)]\,. \tag{65}$$

Now, combining (63) and (65), we obtain

$$\alpha_k^2 + \|x^k - x\|^2 - \|x^{k+1} - x\|^2 \geq$$

$$\frac{2\gamma_k\alpha_k}{\eta_k}\sum_{i=1}^{q(k)}\lambda_i^k(y_i^k)^t[f(x^k)-f(x)]+2\gamma_k\|v^k\|^2+\frac{2\gamma_k\alpha_k}{\eta_k}\max_{y\in G}\left\{y^tJ_f(x^k)v^k\right\}\geq$$

$$\frac{2\gamma_k\alpha_k}{\eta_k}\min_{y\in G}\{[f(x^k)-f(x)]^ty\}+2\gamma_k\|v^k\|^2+\frac{2\gamma_k\alpha_k}{\eta_k}(\bar{y}^k)^tJ_f(x^k)v^k, \tag{66}$$

with $\bar{y}^k$ as in the proof of Proposition 6, using (55) in the last inequality. From (66), (52), Proposition 6 and (25), which implies $\gamma_k\in[0,1]$ for all k, we get

$$\alpha_k^2+\|x^k-x\|^2-\|x^{k+1}-x\|^2\geq\frac{2\gamma_k\alpha_k}{\eta_k}\min_{y\in G}\{[f(x^k)-f(x)]^ty\}-\frac{2\gamma_k\alpha_k}{\eta_k}\|J_f(x^k)^t\bar{y}^k\|\|v^k\|$$

$$\geq\frac{2\gamma_k\alpha_k}{\eta_k}\min_{y\in G}\{[f(x^k)-f(x)]^ty\}-2\gamma_k\alpha_k\|v^k\|\geq\frac{2\gamma_k\alpha_k}{\eta_k}\min_{y\in G}\{[f(x^k)-f(x)]^ty\}-2\alpha_k^2\,, \tag{67}$$

and the result follows immediately from (67). □

We recall next the so called quasi-Fejér convergence theorem.

Proposition 8. *Let $T\subset\mathbb{R}^n$ be a nonempty set and $\{a^k\}\subset\mathbb{R}^n$ a sequence such that $\left\|a^{k+1}-z\right\|^2\leq\left\|a^k-z\right\|+\epsilon_k$ for all $z\in T$ and all k, where $\{\epsilon_k\}\subset\mathbb{R}_+$ is a summable sequence. Then $\{a^k\}$ is bounded. Moreover, if a cluster point $\bar{a}$ of $\{a^k\}$ belongs to T, then the whole sequence $\{a^k\}$ converges to $\bar{a}$.*

Proof. See Theorem 1 in [4]. □

Next we show that, under mild conditions, the sequence $\{x^k\}$ generated by the algorithm is convergent. Let $T=\{x\in C: f(x)\preceq f(x^k)\ \forall k\}$.

Proposition 9. *If $T\neq\emptyset$ then the sequence $\{x^k\}$ generated by the algorithm converges to some $x^*\in T$.*

Proof. Take $x\in T$. Then, for all $y\in K^*$, $[f(x^k)-f(x)]^ty\geq 0$, implying that

$$2\frac{\gamma_k\alpha_k}{\eta_k}\min_{y\in G}\{[f(x^k)-f(x)]^ty\}\geq 0. \tag{68}$$

By (68) and Proposition 7,

$$\left\|x^{k+1}-x\right\|^2\leq\left\|x^k-x\right\|^2+3\alpha_k^2,$$

for all $x\in T$. Since $\{\alpha_k^2\}$ is summable, we apply the first statement of Proposition 8 to conclude that $\{x^k\}$ is bounded. Let $\bar{x}$ be a cluster point of $\{x^k\}$. By (59), $y^t[-J_f(x^k)v^k]\geq 0$ for all $y\in K^*$, so that $J_f(x^k)v^k\preceq 0$. In view of (25) and (26),

$$f(x^{k+1})\preceq f(x^k)+\sigma 2^{-\ell(k)}J_f(x^k)v^k\preceq f(x^k), \tag{69}$$

because $\sigma \geq 0$. By (69) and continuity of f, we get that

$$f(\bar{x}) \preceq f(x^k) \tag{70}$$

for all k, so that $\bar{x} \in T$. By the second statement of Proposition 8, $\{x^k\}$ converges to some $x^* \in T$. □

We close this section with the convergence theorem in the K-convex case.

Theorem 2. *If f is K-convex and $T \neq \emptyset$ then the sequence $\{x^k\}$ generated by the algorithm converges to a weakly efficient solution of problem* (1)–(2).

Proof. By Proposition 9, $\{x^k\}$ converges to some $x^* \in T$. Since $\{x^k\} \subset C$, as commented at the beginning of this section, we have that x^* belongs to C. Suppose that x^* is not weakly efficient. Then, there exists $\hat{x} \in C$ such that $f(\hat{x}) \prec f(x^*)$, implying that $[f(x^*) - f(\hat{x})]^t y > 0$ for all $y \in K^*$. By (70), $\hat{x} \in T$. Also, $\min_{y \in G}\{[f(x^*) - f(\hat{x})]^t y\} > 0$, by compactness of G. Since $\{x^k\}$ is bounded by Propositon 9, we have that $\{J_f(x^k)\}$ is bounded, so that, in view of (52), $\{\eta_k\}$ is bounded. Take ρ such that $\eta_k \leq \rho$ for all k, and let $\mu = \min_{y \in G}\{[f(x^*) - f(\hat{x})]^t y\}$. Using (70) in (62), we get

$$3\alpha_k^2 + \left\|x^k - \hat{x}\right\|^2 - \left\|x^{k+1} - \hat{x}\right\|^2 \geq \frac{2\mu\gamma_k\alpha_k}{\rho} \geq 0. \tag{71}$$

We claim that there exists a subsequence $\{\gamma_{j_k}\}$ of $\{\gamma_k\}$ such that $\lim_{k\to\infty} \gamma_{j_k} = 0$. Summing (71) with k between 0 and p, we get

$$0 \leq \frac{2\mu}{\rho} \sum_{k=0}^{p} \alpha_k\gamma_k \leq 3\sum_{k=0}^{p} \alpha_k^2 + \left\|x^0 - \hat{x}\right\|^2 - \left\|x^{p+1} - \hat{x}\right\|^2 \leq 3\sum_{k=0}^{\infty} \alpha_k^2 + \left\|x^0 - \hat{x}\right\|^2. \tag{72}$$

By (72), since $\rho > 0$, $\mu > 0$, we have

$$\sum_{k=0}^{\infty} \alpha_k\gamma_k \leq \frac{\rho\left(3\sum_{k=0}^{\infty} \alpha_k^2 + \left\|x^0 - \hat{x}\right\|^2\right)}{2\mu} < \infty. \tag{73}$$

If the claim does not hold, then there exists $\bar{\gamma} > 0$ such that $\gamma_k \geq \bar{\gamma}$ for all k, and then, by (73), $\bar{\gamma}\sum_{k=0}^{\infty} \alpha_k < \infty$, in contradiction with $\sum_{k=0}^{\infty} \alpha_k = \infty$. Thus, there exists $\{\gamma_{j_k}\}$ such that

$$\lim_{k\to\infty} \gamma_{j_k} = 0. \tag{74}$$

Let now $\xi_k = \left\|\hat{x} - x^k\right\|^2$, $u^k = \frac{\|v^k\|}{\xi_k}(\hat{x} - x^k)$, so that $\left\|u^k\right\| = \left\|v^k\right\|$. Note that $\lim_{k\to\infty} \alpha_k = 0$, because $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$. Since $\left\|v^k\right\| \leq \alpha_k$ by Proposition 6, we get that $\lim_{k\to\infty} \left\|v^k\right\| = 0$. On the other hand, $\lim_{k\to\infty} \xi_k = \|\hat{x} - x^*\| > 0$, because $f(\hat{x}) \prec f(x^*)$. Thus, $\lim_{k\to\infty} \frac{\|v^k\|}{\xi_k} = 0$, i.e. $\frac{\|v^k\|}{\xi_k} < 1$ for large enough k. Hence,

$$u^k + x^k = \left(1 - \frac{\|v^k\|}{\xi_k}\right) x^k + \frac{\|v^k\|}{\xi_k}\hat{x} \in C,$$

and therefore $u^k \in C_k$ for large enough k. By (51), (53),

$$\psi_k(v^k) \le \psi_k(u^k) = \frac{1}{2}\left\|u^k\right\|^2 + \frac{\alpha_k \left\|v^k\right\|}{\eta_k \xi_k} \max_{y\in G}\{y^t J_f(x^k)(\hat{x} - x^k)\}. \tag{75}$$

Since $\left\|u^k\right\| = \left\|v^k\right\|$, (75) implies

$$\max_{y\in G}\{y^t J_f(x^k) v^k\} \le \frac{\left\|v^k\right\|}{\xi_k} \max_{y\in G}\{y^t J_f(x^k)(\hat{x} - x^k)\}. \tag{76}$$

Refine the subsequence $\{x^{j_k}\}$ so that $\{\left\|v^{j_k}\right\|^{-1} v^{j_k}\}$ converges, say to $\bar{v}$, and let $\nu = \|\hat{x} - x^*\|$. Taking limits as k goes to ∞ along such a subsequence in (76), after dividing both sides by $\left\|v^k\right\|$, we get

$$\max_{y\in G}\{y^t J_f(x^k)\bar{v}\} \le \frac{1}{\nu} \max_{y\in G}\{y^t J_f(x^*)(\hat{x} - x^*)\} \le \frac{1}{\nu} \max_{y\in G}\{y^t[f(\hat{x}) - f(x^*)]\} = -\frac{\mu}{\nu} < 0, \tag{77}$$

using the gradient-like inequality for K-convex functions in the second inequality. It follows from (30), (32) and (77) that

$$J_f(x^*)\bar{v} \prec 0. \tag{78}$$

Let now $s_k = \left\|v^k\right\|$, $\bar{v}^k = \left\|v^k\right\|^{-1} v^k$. Then

$$\frac{1}{s_k}\left[f(x^k) - f(x^k + v^k) + \sigma J_f(x^k) v^k\right] = \frac{f(x^k) - f(x^k + s_k \bar{v}^k)}{s_k} + \sigma J_f(x^k)\bar{v}^k. \tag{79}$$

Since $\lim_{k\to\infty} s_k = \lim_{k\to\infty} \left\|v^k\right\| = 0$, taking limits as k goes to ∞ along the subsequence in (79), we get, in view of the continuous differentiability of f,

$$\lim_{k\to\infty} \frac{1}{s_{j_k}}\left[f(x^{j_k}) - f(x^{j_k} + v^{j_k}) + \sigma J_f(x^{j_k}) v^{j_k}\right] = -J_f(x^*)\bar{v} + \sigma J_f(x^*)\bar{v} = (\sigma - 1) J_f(x^*)\bar{v} \succ 0, \tag{80}$$

using (78), (79) and the fact that $\sigma \in [0, 1)$. It follows from (80) that, for large enough k,

$$\frac{1}{s_{j_k}}\left[f(x^{j_k}) - f(x^{j_k} + v^{j+k}) + \sigma J_f(x^{j_k}) v^{j_k}\right] \succ 0,$$

so that

$$f(x^{j_k} + v^{j_k}) \preceq f(x^{j_k}) + \sigma J_f(x^{j_k}) v^{j_k}. \tag{81}$$

By (81) and (25), $\ell(j_k) = 0$ for large enough k, and henceforth $\gamma_{j_k} = 1$, in contradiction with (74). The contradiction arises from asuming that $f(\hat{x}) \prec f(x^*)$ for some $\hat{x} \in C$, and so the result holds. $\square$

We comment now on the fact that we request both varying α_k's and γ_k's in order to get our convergence result for K-convex objective, which was not needed in the cases we are generalizing. It should be preferable to keep one of such sequences constant, but it seems difficult. If we take, e.g., $\alpha_k = \eta_k = 1$, i.e. we consider the same algorithm for which we proved stationarity of the cluster points (if any) in the nonconvex case, we confront the problem that, even in the scalar-valued case, it is not known how to take advantage of the convexity of f in order to get boundedness of the sequence and uniqueness of the cluster point. The similar results in the unconstrained case use strongly the fact that the chosen direction is minus the gradient, in the scalar case ([4]), and a related one, nemely the one given by (19), in the vector-valued case ([11]), while in the constrained case the resulting direction is just a descent one. This issue deserves further study, starting with the scalar-valued case. Regarding the option of suppressing the Armijo search for the γ_k's, as done in [1] for the scalar-valued case, the obstacle lies in the characterization of the set to which the sequence is quasi-Fejér convergent. In the scalar-valued case, it can be proved that, without the Armijo search, such set consists of the minimizers of f. In the vector-valued case, the reasonable generalization is the set T, but in order to establish that the cluster points of $\{x^k\}$ belong to T, we need to have a monotone sequence of functional values, which we obtain with the Armijo search. Without it, i.e. with a possibly nonmonotone sequence of functional values, there seems to be no way to find a set to which the cluster points are known to belong, in the vector-valued case, and the proof argument breaks.

4 Inexact variants of the method

We will prove in this section that, both in the convex and the nonconvex cases, our convergence results are preserved, when using certain directions which are not the exact solution of (34) or (51).

Let us first consider the nonconvex case. We show next that, given a tolerance $\delta \in (0, 1]$, it is enough to find δ-approximate solutions of (34), i.e. directions $v \in C - x$ such that

$$\varphi(J_f(x)v) + \frac{\hat{\beta}}{2} \|v\|^2 \leq \delta\theta(x)\,, \tag{82}$$

Note that, for nonstationary points, approximate solutions of (34) are K-descent directions. Observe also that, since $\theta(x)$ is the optimal value of problem (34), only the exact solution $v(x)$ satisfies (82) with $\delta = 1$.

We now establish that, for nonstationary points $x \in C$ and for any $\delta \in (0,1]$, all δ-approximate solutions of (34) are contained in a closed ball, whose radius depends continuously on x.

Proposition 10. *Suppose that v is an approximate solution of* (34) *with tolerance $\delta \in (0,1]$. Then $\|v\| \leq \frac{2}{\hat{\beta}} \|J_f(x)\|$.*

Proof. From (82) and Proposition 3, we know that

$$\varphi(J_f(x)v) + \frac{\hat{\beta}}{2} \|v\|^2 \leq \delta\theta(x) \leq 0. \tag{83}$$

By (83) and (30)

$$\|v\|^2 \leq -\frac{2}{\hat{\beta}}\varphi(J_f(x)v) = -\frac{2}{\hat{\beta}}\max_{y\in G}\langle J_f(x)v, y\rangle,$$

and the result follows using Cauchy-Schwartz inequality and the fact that $G \subset \{y \in \mathbb{R}^m : \|y\| = 1\}$. □

For a given tolerance $\delta \in (0,1]$, we consider next the algorithm presented in Subsection 2.2, where v^k is now a δ-approximate solution of (34), instead of the exact solution $v^k = v(x^k)$, as prescribed by (24). We present next the approximate version of Theorem 1.

Theorem 3. *For a given tolerance $\delta \in (0,1]$, every accumulation point, if any, of the sequence $\{x^k\}$ generated by the method in Subsection 2.3, where v^k is now a δ-approximate solution of* (34), *is stationary.*

Proof. Basically the proof is the same as the one of Theorem 1, observing that, for an infinite sequence $\{x^k\}$ (of nonstationary points), we have $J_f(x^k)v^k \prec 0$, so that (42) holds with the inexact direction v^k instead of the exact one. Finally, if $\{x^{j_k}\}$ is the subsequence which converges to the accumulation point, in view of Proposition 10, the sequence of approximate directions $\{v^k\}$ will also be bounded, and the remainder of the proof of Theorem 1 holds for this case without further observations. □

Let us consider now the convex case. The appropriate substitute for the exact direction v^k will be a vector $\hat{v}^k \in C_k$ which is a δ-approximate solution of $\min_{v\in C_k} \psi_k(v)$, and which satisfies also an additional condition. If we consider the scalar-valued function $g_{\hat{y}}(x) = \langle \hat{y}, f(x)\rangle$, with $\hat{y}$ in the convex hull of $G \subset K^*$, we need to have $\hat{v}^k = P_{C_k}\left(-\frac{\alpha_k}{\eta}k\nabla g_{\hat{y}}(x^k)\right)$ for some such $\hat{y}$, where P_{C_k} is the orthogonal projection onto C_k.

At this point two comments are in order: the exact direction v^k is of this form, with $\hat{y} = \sum_{i=1}^{q(k)} \lambda_i^k y_i^k$, as in (55)–(57), as can be easily shown with an argument similar to (21), so that, by continuity of ψ_k, any point y in the convex hull of G close enough to this $\hat{y}$ will have the required form for any given $\delta \in (0,1]$. Also, once such a y is chosen, the $k-$th step of our algorithm is just a step of the scalar projected gradient method applied to minimizing g_y in C. But such a y in general changes with k, because the projected gradient $\hat{v}^k$ is requested to be a δ-approximate minimizer of ψ_k, and our result avoids the already mentioned limitations of a pre-specified scalarization. In fact, both in the convex and nonconvex cases, our algorithm at iteration k chooses some "good" scalarization g_y and performs a projected gradient step on it, "good" meaning that the vector y is close to $\sum_{i=1}^{q(k)} \lambda_i^k y_i^k$ as in (55)–(57). Summarizing, instead of v^k we take a vector $\hat{v}^k \in C_k$ such that:

$$\psi_k(\hat{v}^k) \leq \delta\psi_k(v^k)\,, \tag{84}$$

where $v^k = \mathrm{argmin}_{v\in C_k}\ \psi_k(v)$, and

$$\hat{v}^k := \mathrm{argmin}_{v\in C_k}\ \left\{\frac{1}{2}\|v\|^2 + \frac{\alpha_k}{\eta_k}(\hat{y})^t J_f(x^k)v\right\}\,, \tag{85}$$

for some $\hat{y}$ in the convex hull of G. As mentioned above, an argument as in (21) shows that (85) is equivalent to $\hat{v}^k = P_{C_k}\left(-\frac{\alpha_k}{\eta_k}\nabla g_{\hat{y}}(x^k)\right)$. Also, conditions (84)–(85) extend to the constrained case the notion of δ-approximate scalarization-compatible solution, proposed in [11] for unconstrained vector optimization.

We next show that a vector $\hat{v}^k$ satisfying (84)–(85) basically satisfies the inequality proved in Proposition 6 for $v^k = \mathrm{argmin}_{v\in Ck}\ \psi_k(v)$ and also verifies conditions analogous to (54)–(57).

Proposition 11. *Assume that x^k is an iterate generated by (25), (26) and (51), using, instead of v^k given by (24), a vector $\hat{v}^k \in C_k$ satisfying (84)–(85). Then*

i) $\left\|\hat{v}^k\right\| \leq 2\alpha_k$ *for all* k.

ii) $\langle \hat{v}^k + \frac{\alpha_k}{\eta_k}J_f(x^k)\hat{y}, v - \hat{v}^k\rangle \geq 0$, *for all* $v \in C_k$.

Proof. i) Since $0 \in C_k$, we have $\psi_k(\hat{v}^k) \leq \delta\psi_k(v^k) \leq 0$. From this inequality and the definition of ψ_k, given in (53), we get

$$\|\hat{v}^k\|^2 \leq -\frac{2\alpha_k}{\eta_k}\varphi(J_f(x^k)\hat{v}^k) = -\frac{2\alpha_k}{\eta_k}\mathrm{min}_{y\in G}\left\{(\hat{v}^k)^t J_f(x^k)(y)\right\}. \tag{86}$$

Therefore, from (86), the Cauchy-Schwartz inequality and (52) we obtain

$$\|\hat{v}^k\|^2 \leq 2\frac{\alpha_k}{\eta_k}\left\|J_f(x^k)\bar{y}^k\right\|\left\|\hat{v}^k\right\| \leq 2\alpha_k\left\|\hat{v}^k\right\|, \tag{87}$$

where $\bar{y}^k$ minimizes $\left\{(\hat{v}^k)^t J_f(x^k)y\right\}$ in G, and the result follows immediately from (87).

ii) It suffices to observe that the required inequality is a necessary and sufficient condition for the optimality of $\hat{v}^k$ in (85). ∎

We close this section with the approximate version of Theorem 2.

Theorem 4. *Let* $\{x^k\}$ *be a sequence generated by* (25), (26) *and* (51), *using, instead of* v^k *given by* (24), *a vector* $\hat{v}^k \in C_k$ *satisfying* (84)–(85). *Assume that* f *is* K*-convex and that the set* T*, defined just before Proposition 9, is nonempty. Then* $\{x^k\}$ *converges to a weakly efficient solution of problem* (1)–(2).

Proof. By virtue of Proposition 11, we have that, in this case, all results of Section 3 hold with minor changes. Basically, whenever we have α_k in Section 3, we will now have $2\alpha_k$, and (63) holds for $\hat{v}^k$ by Proposition 11(ii). That is all we need in order to establish the result of Theorem 2 in the new setting. □

5 Final remarks

This paper is part of a series where several well known methods for scalar-valued optimization are extended to the vector-valued case. Here we deal with the projected gradient method, and we make no claims on the advantages and/or drawbacks of this algorithm viz-a-viz others, either in the scalar or in the vector-valued case.

When comparing our approach with the option of applying the projected gradient method to a pre-specified scalarization, we emphasize that it avoids the pitfall of choosing a wrong scalarization, which lacks minimizers, but a price is paid: the minimization (albeit inexact, in the sense of Section 4) of the strongly convex, but nonsmooth function given in (19), needed for selecting the K-descent direction. The computational efficiency of the method will thus depend on the choice of a good procedure for solving this minimization subproblem. We leave this issue as a subject for future research.

References

[1] Alber, Ya.I., Iusem, A.N., Solodov, M.V. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Mathematical Programming* **81** (1998) 23-37.

[2] Bereznyev, V.A., Karmanov, V.G., Tretyakov, A.A. The stabilizing properties of the gradient method. *USSR Computational Mathematics and Mathematical Physics* **26** (1986) 84–85.

[3] Bertsekas, D. *Nonlinear Programming*. Athena Scientific, Belmont (1995).

[4] Burachik, R., Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F. Full convergence of the steepest descent method with inexact line searches. *Optimization* **32** (1995) 137-146.

[5] Correa, R., Lemaréchal, C. Convergence of some algorithms for convex minimization. *Mathematical Programming* **62** (1993) 261–275.

[6] Da Cunha, N.O., Polak, E. Constrained minimization under vector-valued criteria in finite dimensional spaces. *Journal of Mathematical Analysis and Applications* **19** (1967) 103-124.

[7] Fliege, J., Svaiter, B.F. Steepest descent methods for multicriteria optimimization. *Mathematical Methods of Operations Research* **51** (2000) 479-494.

[8] Gafni, E.N., Bertsekas, D. Convergence of a gradient projection method. Technical Report LIDS-P-1201, Laboratory for Information and Decision Systems, M.I.T. (1982).

[9] Geoffrion, A.M. Proper efficiency and the theory of vector maximization. *Journal of Mathematical Analysis and Applications* **22** (1968) 618-630.

[10] Golstein, E., Tretyakov, N. *Modified Lagrange Functions*. Moscow (1989).

[11] Graña Drummond, L.M., Svaiter, B.F. A steepest descent method for vector optimization. Technical Report A2001/21, IMPA, Rio de Janeiro (2001).

[12] Hiriart Urruty, J.-B.; Lemaréchal, C. *Convex Analysis and Minimization Algorithms*. Springer, Berlin (1993).

[13] Iusem, A.N., Svaiter, B.F. A proximal regularization of the steepest descent method. *RAIRO, Recherche Opérationelle* **29**,2 (1995) 123-130.

[14] Kiwiel, K., Murty, K.G. Convergence of the steepest descent method for minimizing quasi convex functions. *Journal of Optimization Theory and Applications* **89** (1996) 221-226.

[15] Luc, D.T. *Theory of Vector Optimization*. Lecture Notes in Economy and Mathematic Systems **319**. Springer, Berlin (1989).

[16] Nesterov, Y.E. *Effective Methods in Nonlinear Programming*. Moscow (1989).

[17] Polyak, B.T. *Introduction to Optimization*. Optimization Software, New York (1987).

[18] Rockafellar, R.T. *Convex Analysis*. Princeton University Press, New Jersey (1970).

[19] Zangwill, W.I. *Nonlinear Programming: a Unified Approach*. Prentice Hall, New Jersey (1969).