Available online at www.sciencedirect.com
SCIENCE DIRECT®

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Journal of Computational and Applied Mathematics 175 (2005) 395–414

www.elsevier.com/locate/cam

A steepest descent method for vector optimization

L.M. Graña Drummond[a,1], B.F. Svaiter[b,*,2]

[a] FACC–UFRJ, Av. Pasteur 250, CEP 22290–240, Rio de Janeiro, RJ, Brazil
[b] Instituto de Matemática Pura e Aplicada (IMPA), Estrada Dona Castorina 110, CEP 22460-320, Rio de Janeiro, RJ, Brazil

Received 24 March 2003; received in revised form 3 June 2004

Abstract

In this work we propose a Cauchy-like method for solving smooth unconstrained vector optimization problems. When the partial order under consideration is the one induced by the nonnegative orthant, we regain the steepest descent method for multicriteria optimization recently proposed by Fliege and Svaiter. We prove that every accumulation point of the generated sequence satisfies a certain first-order necessary condition for optimality, which extends to the vector case the well known "gradient equal zero" condition for real-valued minimization. Finally, under some reasonable additional hypotheses, we prove (global) convergence to a weak unconstrained minimizer.

As a by-product, we show that the problem of finding a weak constrained minimizer can be viewed as a particular case of the so-called Abstract Equilibrium problem.

Keywords: Pareto optimality; Vector optimization; Steepest descent; K-convexity; Quasi-Féjer convergence

1. Introduction

In multicriteria optimization, several objective functions have to be minimized simultaneously. Usually, no single point will minimize all given objective functions at once (i.e., there does not exist an *ideal* minimizer), and so the concept of optimality has to be replaced by the concept of *Pareto-optimality* or *efficiency*. A point is called Pareto-optimal or efficient, if there does not exist a different point with

* Corresponding author. Tel.: +55-21-2529-5112; fax: +55-21-2529-5129.
E-mail addresses: lm@ondaalta.com.br (L.M. Graña Drummond), benar@impa.br (B.F. Svaiter).
[1] Partially supported by FAPERJ under Grant E-26/152.107/1999-Bolsa.
[2] Partially supported by CNPq Grant 302748/2002-4, PRONEX-Optimization and FAPERJ.

0377-0427/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2004.06.018

smaller than or equal objective function values, such that there is a decrease in at least one objective function value. Applications for this type of problem can be found in engineering design [12] (mainly truss optimization [8]), location science [5], statistics [6], management science [13] (specially portfolio analysis [24]), etc.

Among the main solution strategies for multicriteria optimization problems, we mention the scalarization approaches [15,16,19,21,23]. Here, one or several parameterized single-objective (i.e., classical) optimization problems are solved. Frequently, some parameters have to be specified in advance, leaving the modeler and the decision-maker with the burden of choosing them. Moreover, in the weighting method, for example, bad choices of these parameters can lead to unbounded scalar problems. Other scalarization techniques are parameter-free [7,8,21] but try to compute a discrete approximation to the whole set of Pareto-optimal points.

Parameter-free multicriteria optimization techniques use in general an ordering of the different criteria, i.e., an ordering of importance of the components of the objective function vector. In this case, the ordering has to be specified. Moreover, the optimization process is usually augmented by an interactive procedure [20], adding an additional burden to the task of the decision-maker.

In a recent paper, Fliege and Svaiter [14] proposed a Pareto descent method for multiobjective optimization. This procedure is parameter-free and relies upon a suitable extension for vector-valued functions of the classical steepest descent direction. Neither ordering information nor weighting factors for the different objective functions is assumed to be known in this new method, which may be interpreted as a "Cauchy's method" for multicriteria optimization.

We recall that the steepest descent method (also known as gradient or Cauchy's method) is one of the oldest and more basic minimization schemes for scalar unconstrained optimization. Despite its computational shortcomings, like, for instance, "hemstitching" phenomena, the Cauchy's method can be considered among the most important procedures for minimization of real-valued functions defined on $\mathbb{R}^n$, since it is the departure point for many other more sophisticated and efficient algorithms. For instance, it is partially used in some "globally convergent" modifications of Newton's method for unconstrained optimization. Here, "globally convergent" means that all sequences produced by the method have decreasing objective function values, and that all accumulation points of these sequences are critical points. We refer the reader to [9], where the "double dog-leg" method is discussed. The simple idea of decreasing the value of the objective function is also used in many other modifications of Newton's method. We refer the reader again to [9] for a very clear exposition. It remains an open question how to extend more efficient procedures, as Newton's method, to vector optimization.

The purpose of this paper is to take a step further on the direction of Fliege and Svaiter's [14] work. Based on their ideas, we present a Cauchy-like method for smooth vector optimization. In this setting, the partial order is induced by a general closed convex pointed cone K, with nonempty interior in a finite-dimensional space. Our procedure depends on the choice of an arbitrary initial point, as well as on a certain compact set which characterizes the positive polar cone (see Section 3). On the final remarks we make more comments on this issue. When the cone is the nonnegative orthant, our procedure turns out to be the very same proposed by them for the unconstrained case. Our convergence results extend theirs and, furthermore, under some additional (and quite reasonable) hypotheses, we also show that all sequences produced by the method converge to a weakly efficient point, no matter how poor is the initial guess. We point out that we are not attempting to find the set of all efficient or weak efficient optima.

Regarding the importance of vector optimization, we point out that even though the vast majority of real life problems formulated as vector-valued problems deals with the component-wise partial order, i.e., the one which arises from the Paretian cone, there are many others that require preference orders induced by closed convex cones other than the nonnegative orthant. Such cones have been recently analyzed, for example in [1] (based on the theoretical results of [2]), where problems of portfolio selection in security markets require finding weak Pareto minimal points with respect to feasible portfolio cones, which are nonlattice, that is to say cones with more extreme rays than the ambience space dimension.

The outline of this work is as follows. In Section 2, we introduce the unconstrained K-minimization problem and discuss a necessary (but in general nonsufficient) first-order optimality condition, called K-criticality. This condition extends to vector optimization the classical "gradient equal zero" condition for scalar minimization. The notion of K-criticality allows us to derive a generic K-descent scheme. We discuss an Armijo-like strategy for choosing the stepsizes. In Section 3, we characterize the negative cone and its interior in terms of a convex function. This function is used to define the steepest descent direction, as well as its approximations. We present the complete steepest descent method for vector optimization, with approximations of the steepest descent direction, and show its well definedness. In Section 4, we discuss the convergence of the method. We prove that all the cluster points of any sequence generated by the algorithm are K-critical. In Section 5, we discuss the relationship between vector steepest descent direction and the scalarization approach. In Section 6, assuming K-convexity of the objective function and a certain very reasonable condition on the objective, we prove, based upon the notion of quasi-Fejér convergence, that, with the pure steepest descent direction choice or even with other more general directions, called s-compatibles, the method is globally convergent to a weak unconstrained K-minimizer. In Section 7, using the scalar function defined in Section 3, we show that the problem of finding a weak constrained K-minimizer of a vector-valued function can be viewed as a particular case of the well known abstract equilibrium problem. Finally, in Section 8 we make some final remarks about our work.

2. Basic definitions

Let K be a closed pointed convex cone of $\mathbb{R}^m$, with nonempty interior. The partial order in $\mathbb{R}^m$ induced by K, $\preccurlyeq_K$, is defined by:

$$u \preccurlyeq_K v \quad \text{if } v - u \in K.$$

Consider also the following relation induced by $\text{int}(K)$ in $\mathbb{R}^m$, $\prec_K$:

$$u \prec_K v \quad \text{if } v - u \in \text{int}(K).$$

Given a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$, we consider the problem of finding an *unconstrained K-minimizer* (or *K-optimum*) of F, i.e., a point $x^* \in \mathbb{R}^n$ such that there exists no other $x \in \mathbb{R}^n$ with $F(x) \preccurlyeq_K F(x^*)$ and $F(x) \neq F(x^*)$. In other words, we are seeking unconstrained minimizers for F in the partial order induced by the cone K. We denote this problem as

$$\min_K F(x). \tag{1}$$

A necessary, but, in general, nonsufficient, condition for K-optimality of a point $x \in \mathbb{R}^n$ is

$$-\text{int}(K) \cap \text{Image}(JF(x)) = \emptyset, \tag{2}$$

where $JF(x)$ stands for the Jacobian of F at x and $\mathrm{Image}(JF(x))$ stands for the image of $\mathbb{R}^n$ by the linear operator $JF(x)$ (see, for instance, [19]). Observe that (2) generalizes to vector optimization the classical condition "gradient equal zero" for the real-valued case.

A point x is *K-critical* if it satisfies (2). Therefore, if a point x is not K-critical, there exists a direction $v \in \mathbb{R}^n$ satisfying

$$JF(x)v \in -\mathrm{int}(K)\,, \tag{3}$$

that is to say $JF(x)v \prec_K 0$. It is a well known fact that such v is a *K-descent direction* for the objective F. Actually, it holds that (see [19]), if v satisfies (3), there exists $\bar{t} > 0$ such that,

$$F(x+tv) \prec_K F(x) \quad \text{for all } t \in (0, \bar{t}). \tag{4}$$

If v is a K-descent direction at x, we say that $t > 0$ satisfies the "Armijo-like" rule for $\beta \in (0,1)$ if

$$F(x+tv) \preccurlyeq_K F(x) + \beta t JF(x)v.$$

As in the scalar case, given a "descent" direction, the Armijo rule is satisfied for some t's.

Proposition 2.1. *Let $\beta \in (0,1)$. If $JF(x)v \prec_K 0$, then there exists $\bar{t} > 0$ such that,*

$$F(x+tv) \prec_K F(x) + \beta t JF(x)v$$

for all $t \in (0, \bar{t})$.

Proof. Since F is differentiable, we have

$$F(x+tv) = F(x) + t(JF(x)v + R(t)), \tag{5}$$

with $\lim_{t \to 0} R(t) = 0$. We are assuming that $JF(x)v \in -\mathrm{int}(K)$. Since $\beta \in (0,1)$, $(1-\beta)JF(x)v \in -\mathrm{int}(K)$. Hence, there exists $\bar{t} > 0$ such that, for all $t \in (0, \bar{t}]$, $\|R(t)\|$ is small enough, so that

$$R(t) + (1-\beta)JF(x)v \in -\mathrm{int}(K).$$

Equivalently,

$$R(t) \prec_K -(1-\beta)JF(x)v \quad \text{for all } t \in (0, \bar{t}].$$

Combining this K-inequality with (5) the conclusion follows. □

Proposition 2.1 opens the way for defining a generic K-descent method using the Armijo-like rule.

Algorithm 1 (*Generic K-descent method*).

1. *Take $\beta \in (0,1)$, $x^0 \in \mathbb{R}^n$. Set $k := 0$.*
2. *If x^k is K-critical* STOP. *Otherwise,*
3. *Find v^k a K-descent direction at x^k.*

4. *Find* $t_k > 0$ *such that*

$$F(x^k + t_k v^k) \preccurlyeq_K F(x^k) + \beta t_k JF(x^k)v^k.$$

5. *Set* $x^{k+1} := x^k + t_k v^k$, $k := k+1$ *and* GOTO 2.

Only very general properties can be proved for this algorithm.

Proposition 2.2. *Let* $\{x^k\}$ *be an infinite sequence generated by Algorithm* 1. *If* $\bar{x}$ *is an accumulation point of* $\{x^k\}$ *then*

$$F(\bar{x}) \preccurlyeq_K F(x^k)$$

for all k and $\lim_{k\to\infty} F(x^k) = F(\bar{x})$. *In particular, F is constant in the set of accumulation points of* $\{x^k\}$.

Proof. We are supposing that an infinite sequence $\{x^k\}$ was generated by Algorithm 1. Therefore, all x^k are non K-critical and $\{F(x^k)\}$ is K-decreasing. By assumption, there is a subsequence $\{x^{k_j}\}$ converging to $\bar{x}$. Take any $k \in \mathbb{N}$. For j large enough $k_j > k$ and

$$F(x^{k_j}) \preccurlyeq_K F(x^k).$$

Taking the limit for $j \to \infty$ we get $F(\bar{x}) \preccurlyeq_K F(x^k)$. Let $\hat{x}$ be another accumulation point of $\{x^k\}$. Then there exists a subsequence $\{x^{k_p}\}$ converging to $\hat{x}$. Since

$$F(\bar{x}) \preccurlyeq_K F(x^{k_p}),$$

letting $p \to \infty$ we get $F(\bar{x}) \preccurlyeq_K F(\hat{x})$. By the same reasoning, $F(\hat{x}) \preccurlyeq_K F(\bar{x})$. Since K is pointed, these two K-inequalities imply that $F(\hat{x}) = F(\bar{x})$. □

Regarding the choice of the stepsizes t_k, if they are taken too small, the generated sequence $\{x^k\}$ may converge to a non K-critical point. In order to choose a suitable steplength at iteration k, we prescribe the usual backtracking procedure:

3a. Set $t := 1$
3b. If $F(x^k + tv^k) \preccurlyeq_K F(x^k) + \beta t JF(x^k)v^k$ then

$$t_k = t, \ \text{END (of backtracking)}$$

else
3c. Set $t := t/2$, GOTO 3b.

Observe that the above backtracking procedure has always finite termination, thanks to Proposition 2.1. Moreover

$$t_k = \max\{2^{-j} \mid j \in \mathbb{N},\ F(x^k + 2^{-j}v^k) \preccurlyeq_K F(x^k) + \beta 2^{-j} JF(x^k)v^k\}.$$

The main problem now is the choice of v^k. For classical optimization, we have $m = 1$, $K = \mathbb{R}_+$. In this case, the natural choice is the steepest descent direction $v^k = -\nabla F(x^k)$, which happens to be the solution of

$$\min \quad \langle v, \nabla F(x^k)\rangle + (1/2)\|v\|^2, \quad v \in \mathbb{R}^n. \tag{6}$$

For multiobjective optimization ($m \geqslant 1$), where K is the positive orthant $\mathbb{R}^m_+$, Fliege and Svaiter [14] proposed to take v^k as the solution of

$$\min \quad \max_{i=1,\ldots,m} \langle v, \nabla F_i(x^k)\rangle + (1/2)\|v\|^2, \quad v \in \mathbb{R}^n, \tag{7}$$

where $F(x) = (F_1(x), \ldots, F_m(x))$. Observe that (6) is a particular case of (7) when $m = 1$.

In the following section, we will extend the notion of steepest descent direction for the partial order $\preccurlyeq_K$ and propose the K-steepest descent method.

3. K-steepest descent method

The positive polar cone of $A \subseteq \mathbb{R}^m$ is the set

$$A^* = \{w \in \mathbb{R}^m \mid \langle y, w\rangle \geqslant 0 \quad \forall y \in A\}.$$

Since K is a closed convex cone, $K = K^{**}$ (see [22, Theorem 14.1]) and

$$-K = \{y \in \mathbb{R}^m \mid \langle y, w\rangle \leqslant 0 \quad \forall w \in K^*\}$$

and

$$-\mathrm{int}(K) = \{y \in \mathbb{R}^m \mid \langle y, w\rangle < 0 \quad \forall w \in K^* \backslash \{0\}\}.$$

The convex hull of $A \subseteq \mathbb{R}^m$ will be denoted by $\mathrm{conv}(A)$, and the cone generated by A will be denoted by $\mathrm{cone}(A)$.

From now on, we assume that we have a compact set $C \subseteq \mathbb{R}^m$ such that:

$$0 \notin C, \tag{8}$$

$$\mathrm{cone}(\mathrm{conv} C) = K^*. \tag{9}$$

As $\mathrm{int}(K) \neq \emptyset$ and $C \subseteq K^* \backslash \{0\}$, it follows that $0 \notin \mathrm{conv}(C)$. Therefore

$$-K = \{u \in \mathbb{R}^m \mid \langle u, w\rangle \leqslant 0 \quad \forall w \in C\}, \tag{10}$$

$$-\mathrm{int}(K) = \{u \in \mathbb{R}^m \mid \langle u, w\rangle < 0 \quad \forall w \in C\}. \tag{11}$$

In classical optimization, $K = \mathbb{R}_+$ and we may take $C = \{1\}$. For multiobjective optimization, K and K^* are the positive orthant of $\mathbb{R}^m$ and we may take C as the canonical basis of $\mathbb{R}^m$. If K is a polyhedral cone, C may be taken as a finite set of extremal rays of K^*. For a generic K (closed pointed convex cone with nonempty interior), the set

$$C = \{w \in K^* \mid \|w\|_1 = 1\}$$

(where $\|w\|_1 = |w_1| + \cdots + |w_m|$) will satisfy conditions (8), (9).

Define now $\varphi : \mathbb{R}^m \to \mathbb{R}$,

$$\varphi(y) := \sup_{w \in C} \langle y, w\rangle. \tag{12}$$

In view of (10)–(11) and the compactness of C, the function φ gives a "scalar" characterization of $-K$ and $-\mathrm{int}(K)$:

$$-K = \{y \in \mathbb{R}^m \mid \varphi(y) \leqslant 0\} \tag{13}$$

and

$$-\mathrm{int}(K) = \{y \in \mathbb{R}^m \mid \varphi(y) < 0\}. \tag{14}$$

In the following lemma we establish some elementary properties of the function φ which will be used in the sequel.

Lemma 3.1. (i) *Let* $y, y' \in \mathbb{R}^m$, *then* $\varphi(y + y') \leqslant \varphi(y) + \varphi(y')$ *and* $\varphi(y) - \varphi(y') \leqslant \varphi(y - y')$. (ii) *Let* $y, y' \in \mathbb{R}^m$, *if* $y \prec_K y'$ $(y \preccurlyeq_K y')$, *then* $\varphi(y) < \varphi(y')$ $(\varphi(y) \leqslant \varphi(y'))$. (iii) *The function* $\varphi : \mathbb{R}^n \to \mathbb{R}$ *is Lipschitz continuous.*

Proof. Item (i). The first inequality holds trivially and the second follows from the first. Item (ii). The fact that $y - y' \in -\mathrm{int}(K)$ $(y - y' \in -K)$ is equivalent to $\varphi(y - y') < 0$ $(\varphi(y - y') \leqslant 0)$, according to (14) and (13). Hence, the result follows from (i).

Item (iii). By virtue of what was established in (i), $\varphi(y) - \varphi(y') \leqslant \varphi(y - y')$ and $\varphi(y') - \varphi(y) \leqslant \varphi(y' - y)$. Hence

$$|\varphi(y) - \varphi(y')| \leqslant \sup\{\varphi(y - y'), \varphi(y' - y)\}.$$

Therefore, from (12) and Cauchy–Schwartz inequality,

$$|\varphi(y) - \varphi(y')| \leqslant L\|y - y'\|, \tag{15}$$

where, $L := \sup\{\|w\| \mid w \in C\}$, and the result follows. □

Define now for $x \in \mathbb{R}^n$, $f_x : \mathbb{R}^n \to \mathbb{R}$ as

$$\begin{aligned} f_x(v) &:= \varphi(JF(x)v) \\ &= \sup_{w \in C} \langle w, JF(x)v \rangle. \end{aligned} \tag{16}$$

From (14) it follows that v is a K-descent direction at x, if and only if, $f_x(v) < 0$. Therefore, x is K critical if and only if $f_x(v) \geqslant 0$ for all $v \in \mathbb{R}^n$.

We can now extend the notion of steepest descent direction to the vector case (with K an arbitrary cone satisfying the conditions stated at the beginning of Section 2).

Definition 3.2. Given $x \in \mathbb{R}^n$, the K-steepest descent direction (for F) at x, denoted by v_x is the solution of

$$\min \; f_x(v) + (1/2)\|v\|^2, \quad v \in \mathbb{R}^n. \tag{17}$$

The optimal value of this problem will be denoted by α_x.

Remark 1. In the scalar minimization case, where $F : \mathbb{R}^n \to \mathbb{R}$ and $K = \mathbb{R}_+$, taking $C = \{1\}$, the K-steepest descent direction is exactly the classical steepest descent direction, i.e., $v_x = -\nabla F(x)$.

Remark 2. For multicriteria optimization, where $K=\mathbb{R}^m_+$, with C given by the canonical basis of $\mathbb{R}^m$, we retrieve the steepest descent direction proposed in [14].

Since $v \mapsto f_x(v)$ is a real-valued closed convex function, v_x and α_x are well defined. Furthermore, as F is continuously differentiable and φ is Lipschitz continuous, the mapping $(x,v) \mapsto f_x(v)$ is also continuous.

Lemma 3.3.

1. *If x is K-critical then* $v_x=0$, $\alpha_x=0$.
2. *If x is not K-critical then* $v_x \neq 0$, $\alpha_x<0$,

$$f_x(v_x) < -(1/2)\|v_x\|^2 < 0$$

and v_x is a K-descent direction.

3. *The mappings $x \mapsto v_x$, $x \mapsto \alpha_x$ are continuous.*

Proof. Item 1. If x is K-critical, then $f_x(v) \geqslant 0\ \forall v \in \mathbb{R}^n$. Since $f_x(0)=0$, the conclusion follows.

Item 2. If x is not K-critical, then, for some $v \in \mathbb{R}^n$, $f_x(v)<0$. Observe that $f_x(\cdot)$ is positive homogeneous of degree 1. Taking

$$\tilde{t} = -f_x(v)/\|v\|^2, \quad \tilde{v}=\tilde{t}v,$$

we get

$$\begin{aligned} f_x(\tilde{v}) + (1/2)\|\tilde{v}\|^2 &= \tilde{t} f_x(v) + (1/2)\tilde{t}^2\|v\|^2 \\ &= -(1/2) f_x(v)^2/\|v\|^2 < 0. \end{aligned}$$

Hence $\alpha_x<0$. The other statements of item 2 now follow trivially.

Item 3. Take $x^0 \in \mathbb{R}^n$ and $\varepsilon>0$. Define

$$S := \{v \in \mathbb{R}^n \mid \|v_{x^0} - v\| = \varepsilon\}.$$

Note that v_{x^0} is optimal for (17) with $x=x^0$. From (16) it follows that $f_x(\cdot)$ is convex, so the objective function on the minimization problem (17) is strongly convex with modulus $1/2$. It follows that,

$$f_{x^0}(v) + (1/2)\|v\|^2 \geqslant f_{x^0}(v_{x^0}) + (1/2)\|v_{x^0}\|^2 + (1/2)\varepsilon^2 \quad \forall v \in S.$$

Since the mapping $(x,v) \mapsto f_x(v)$ is continuous, and S is compact, using this equation we conclude that there exists $\delta>0$ such that, if $\|x-x^0\| \leqslant \delta$, then

$$f_x(v) + (1/2)\|v\|^2 > f_x(v_{x^0}) + (1/2)\|v_{x^0}\|^2 \quad \forall v \in S.$$

Take now $x \in \mathbb{R}^n$, $\|x-x^0\| \leqslant \delta$. As $v \mapsto f_x(v) + (1/2)\|v\|^2$ is convex, we conclude from the above inequality that v_x, the minimizer of $f_x(\cdot)+(1/2)\|\cdot\|^2$ is not in the region $\|v-v_{x^0}\| \geqslant \varepsilon$, hence $\|v_x - v_{x^0}\| \leqslant \varepsilon$. Continuity of α_x follows now trivially. $\square$

A possible choice for v^k in Algorithm 1 is v_{x^k}, i.e., the K-steepest descent direction at x^k. Since the computation of v_x requires the solution of (17), it would be interesting to work with approximated solutions of this problem.

Definition 3.4. Let $\sigma \in [0, 1)$. We say that v is a σ-approximate K-steepest descent direction at $x \in \mathbb{R}^n$ if

$$f_x(v) + (1/2)\|v\|^2 \leqslant (1-\sigma)\alpha_x,$$

or equivalently

$$f_x(v) + (1/2)\|v\|^2 - \left(f_x(v_x) + (1/2)\|v_x\|^2\right) \leqslant \sigma|\alpha_x|.$$

Observe that the (exact) K-steepest descent direction at x is always a σ-approximate K-steepest descent direction, because we assume $\sigma \in [0, 1)$. The *exact* K-steepest descent direction at x is the unique $\sigma = 0$-approximate K-steepest descent direction.

Lemma 3.5. *Let $\sigma \in [0, 1)$. If v is a σ-approximate K-steepest descent direction at x, then*

$$\|v_x - v\|^2 \leqslant 2\sigma|\alpha_x|.$$

Proof. The function

$$v \mapsto f_x(v) + (1/2)\|v\|^2$$

is strongly convex with modulus $1/2$. Since v_x is the minimizer of this function,

$$f_x(v) + (1/2)\|v\|^2 - \left(f_x(v_x) + (1/2)\|v_x\|^2\right) \geqslant (1/2)\|v_x - v\|^2.$$

Using Definition 3.4, the conclusion follows. $\square$

Let $\sigma \in [0, 1)$ be a prespecified tolerance. From Lemmas 3.3 and 3.5, it follows that $v = 0$ is a σ-approximate K-steepest descent direction at x if, and only if, x is K-critical. Note also that, if x is not K-critical and v is a σ-approximate K-steepest descent direction at x, then v is a K-descent direction, and in particular $v \neq 0$.

Now we formally state the K-steepest descent method (with K-Armijo rule, implemented with backtracking). This algorithm is a particular case of Algorithm 1.

Algorithm 2 (*K-steepest descent method*).

1. *Choose $\beta \in (0, 1)$, $\sigma \in [0, 1)$, $x^0 \in \mathbb{R}^n$. Set $k := 0$.*
2. *If x^k is K-critical (i.e., if $f_{x^k}(v) \geqslant 0$ for all $v \in \mathbb{R}^n$)* STOP. *Otherwise,*
3. *Compute v^k, a σ-approximate K-steepest descent direction at x^k.*
4. *Compute the steplength $t_k \in (0, 1]$ in the following way*:

$$t_k := \max\{2^{-j} \mid j \in \mathbb{N},\ \varphi(F(x^k + 2^{-j}v^k) - F(x^k) - \beta 2^{-j} JF(x^k)v^k) \leqslant 0\}.$$

5. *Set $x^{k+1} := x^k + t_k v^k$, $k := k+1$ and* GOTO 2.

Observe that if x^k is not K-critical, then v^k obtained in step 3 is a K-descent direction and t_k in step 4 is well defined. Moreover, such t_k may be obtained by a backtracking procedure, as discussed previously. Note that, by virtue of (13), in step 4 we have,

$$t_k := \max\{2^{-j} \mid j \in \mathbb{N},\ F(x^k + 2^{-j}v^k) \preccurlyeq_K F(x^k) + \beta 2^{-j} JF(x^k)v^k\}. \tag{18}$$

Furthermore, x^{k+1} will satisfy $F(x^{k+1}) \preccurlyeq_K F(x^k)$. So, the objective values sequence $\{F(x^k)\}$ is K-nonincreasing.

We finish this section with a generalization of Proposition 2.1, which is a simple consequence of the fact that F is continuously differentiable.

Proposition 3.6. *Let $\beta \in (0,1)$, x and v such that $JF(x)v \prec_K 0$. Then there exist $\hat{t}, \delta, \delta' > 0$ such that,*

$$v' \text{ is a } K\text{-descent direction at } x',$$
$$F(x' + tv') \prec_K F(x') + \beta t JF(x')v'$$

for any $t \in (0, \hat{t})$, $x' \in B(x;\delta)$, $v' \in B(v;\delta')$.

Proof. By assumption, $JF(x)v \in -\mathrm{int}(K)$. So, there exists $\varepsilon > 0$ such that

$$JF(x)v + y \in -\mathrm{int}(K) \quad \forall y \in \mathbb{R}^m, \|y\| \leqslant \varepsilon. \tag{19}$$

Since *JF* is continuous, there exist $\delta_1, \delta_2 > 0$ such that if $\|x' - x\| \leqslant \delta_1$, $\|v' - v\| \leqslant \delta_2$, then

$$\|JF(x')v' - JF(x)v\| \leqslant \varepsilon/2 \tag{20}$$

and so

$$JF(x')v' \prec_K 0.$$

Continuity of *JF* also implies that

$$F(z + tu) = F(z) + tJF(z)u + tR(z, tu),$$

with $\lim_{t \to 0} \|R(z, tu)\| = 0$ *uniformly* for z and u in compact sets. Therefore, there exists $\hat{t} > 0$ such that, for $t \in (0, \hat{t})$, $\|x' - x\| \leqslant \delta_1$, $\|v' - v\| \leqslant \delta_2$,

$$\|R(x', tv')\| \leqslant (\varepsilon/2)(1 - \beta). \tag{21}$$

Now, assume that $t \in (0, \hat{t})$, $\|x' - x\| \leqslant \delta_1$, $\|v' - v\| \leqslant \delta_2$. Then

$$\begin{aligned} F(x' + tv') &= F(x') + tJF(x')v' + tR(x', tv') \\ &= F(x') + t\beta JF(x')v' + t[(1 - \beta)JF(x')v' + R(x', tv')]. \end{aligned}$$

Defining

$$u := JF(x')v' + (1 - \beta)^{-1} R(x', tv'), \tag{22}$$

we have

$$F(x' + tv') = F(x') + t\beta JF(x')v' + t(1 - \beta)u. \tag{23}$$

It suffices to see that $u \prec_K 0$. Observe that

$$u = JF(x)v + \tilde{y}, \tag{24}$$

where $\tilde{y} := JF(x')v' - JF(x)v + (1-\beta)^{-1}R(x', tv')$. Using (20) and (21),

$$\begin{aligned}\|\tilde{y}\| &\leqslant \|JF(x')v' - JF(x)v\| + (1-\beta)^{-1}\|R(x', tv')\| \\ &\leqslant \varepsilon.\end{aligned}$$

Therefore, using (19) and (24) we conclude that $u \prec_k 0$ and so, the result follows from (23). □

4. Convergence analysis: the general case

From now on, $\{x^k\}$, $\{v^k\}$, $\{t_k\}$ are sequences generated by Algorithm 2. If the algorithm terminates after a finite number of iterations, it terminates at a K-critical point. In this section, we suppose that an infinite sequence $\{x^k\}$ is generated. So, in view of Lemma 3.3, Definition 3.4 and Proposition 3.6, for all k,

$$\begin{aligned}&\alpha_{x^k} < 0, \\ &f_{x^k}(v^k) + (1/2)\|v^k\|^2 \leqslant (1-\sigma)\alpha_{x^k} < 0, \\ &F(x^{k+1}) \preccurlyeq_K F(x^k) + \beta t_k JF(x^k)v^k \preccurlyeq_K F(x^k).\end{aligned}$$

In particular the sequence $\{F(x^k)\}$ is K-decreasing.

Using the above K-inequality, Lemma 3.1 and the positive homogeneity of φ, it follows that for all k,

$$\begin{aligned}\varphi(F(x^{k+1})) &\leqslant \varphi(F(x^k) + \beta t_k JF(x^k)v^k) \\ &\leqslant \varphi(F(x^k)) + \varphi(\beta t_k JF(x^k)v^k) \\ &= \varphi(F(x^k)) + \beta t_k \varphi(JF(x^k)v^k) \\ &= \varphi(F(x^k)) + \beta t_k f_{x^k}(v^k) \\ &\leqslant \varphi(F(x^k)) + \beta t_k \left((1-\sigma)\alpha_{x^k} - (1/2)\|v^k\|^2\right).\end{aligned} \tag{25}$$

As a consequence of this scalar inequality we obtain the following lemma.

Lemma 4.1. *If $\{F(x^k)\}$ is K-bounded from below, (i.e., if there exists $\bar{y}$ such that $\bar{y} \preccurlyeq_K F(x^k)$ for all k) then,*

$$\sum t_k|\alpha_{x^k}| < \infty, \quad \sum t_k\|v^k\|^2 < \infty.$$

Proof. Adding inequality (25) from $k=0$ to n we get,

$$\begin{aligned}\varphi(F(x^{n+1})) &\leqslant \varphi(F(x^0)) + \sum_{k=0}^{n} \beta t_k((1-\sigma)\alpha_{x^k} - (1/2)\|v^k\|^2) \\ &= \varphi(F(x^0)) - \sum_{k=0}^{n} \beta t_k \left((1-\sigma)|\alpha_{x^k}| + (1/2)\|v^k\|^2\right).\end{aligned}$$

If $\bar{y} \preccurlyeq_K F(x^k)$ for all k, then $\varphi(\bar{y}) \leqslant \varphi(F(x^k))$ for all k and the conclusion follows. □

Now we are in conditions of studying the convergence properties of Algorithm 2.

Theorem 4.2. *All accumulation points of* $\{x^k\}$ *are K-critical.*

Proof. Let $\bar{x}$ be an accumulation point of $\{x^k\}$. Then there exists a subsequence $\{x^{k_j}\}$ converging to $\bar{x}$,

$$\lim_{j\to\infty} x^{k_j} = \bar{x}.$$

Note that $\{v_{x^{k_j}}\}$ and $\{\alpha_{x^{k_j}}\}$ are bounded because they converge to $v_{\bar{x}}$ and $\alpha_{\bar{x}}$, respectively. Therefore, using Lemma 3.5 we conclude that $\{v^{k_j}\}$ is also bounded. So (refining the original sequence if necessary), we may also assume that $\{v^{k_j}\}$ converges to some $\bar{v}$,

$$\lim_{j\to\infty} v^{k_j} = \bar{v}.$$

For all k, $f_{x^k}(v^k) + (1/2)\|v^k\|^2 \leqslant (1-\sigma)\alpha_{x^k}$. Taking limits along $k = k_j$ for $j \to \infty$ we get,

$$f_{\bar{x}}(\bar{v}) + (1/2)\|\bar{v}\|^2 \leqslant (1-\sigma)\alpha_{\bar{x}}. \tag{26}$$

Recall that Algorithm 2 is a particular case of Algorithm 1, so by Proposition 2.2, $F(\bar{x})$ is a K-lower bound for $\{F(x^k)\}$. Now we may apply Lemma 4.1 to conclude that,

$$\lim_{j\to\infty} t_{k_j}\alpha_{x^{k_j}} = 0, \tag{27}$$

$$\lim_{j\to\infty} t_{k_j}\|v^{k_j}\| = 0. \tag{28}$$

We claim that

$$\bar{v} = 0. \tag{29}$$

Suppose, for contradictory purposes, that $\bar{v} \neq 0$. As $\alpha_{\bar{x}} \leqslant 0$, using (26) we get $f_{\bar{x}}(\bar{v}) < 0$. Using Proposition 3.6 we conclude that there exists $\hat{t} > 0$ such that, for j large enough (greater than some j_0)

$$F(x^{k_j} + tv^{k_j}) \preccurlyeq_K F(x^{k_j}) + \beta t JF(x^{k_j})v^{k_j} \quad \forall t \in [0, \hat{t}). \tag{30}$$

Now we will show that for j larger than such j_0,

$$2t_{k_j} \geqslant \min\{1, \hat{t}\}. \tag{31}$$

Indeed, let $j > j_0$. If $t_{k_j} = 1$, the claim holds. If $t_{k_j} < 1$, then, this stepsize was obtained by a backtracking procedure, where the "previous" possible stepsize $2t_{k_j}$ does not satisfy the descend condition, or equivalently, $2t_{k_j}$ does not belong to the k_jth set in (18). Now using (30) we conclude that $2t_{k_j} \geqslant \hat{t}$ and the claim holds also in this case.

The assumption $\bar{v} \neq 0$ also implies, by (28), that $\lim_{j\to\infty} t_{k_j} = 0$, which contradicts (31).

To end the proof, use (26) and (29) to obtain $\alpha_{\bar{x}} \geqslant 0$. Since $\alpha_x \leqslant 0$ for any x, we conclude that $\alpha_{\bar{x}} = 0$ and $\bar{x}$ is K-critical. □

5. Scalarization and Algorithm 2

A very useful method for solving problem (1) is the so-called scalarization procedure. The method is quite elegant; it consists of minimizing a certain scalar function, as explained in the sequel. Take some

$w \in \mathrm{int}(K^*)$ and define $g : \mathbb{R}^n \to \mathbb{R}$,

$$g(x) = \langle w, F(x) \rangle. \tag{32}$$

Then, solutions of

$$\min \ g(x), \quad x \in \mathbb{R}^n$$

are also solutions of (1). So, we only need to minimize a (smooth) scalar function, and for this problem there are many efficient algorithms. The choice of $w \in \mathrm{int}(K^*)$ is of capital importance. Indeed, for very well behaved problems, many choices of w lead to unbounded scalar minimizations problems. For example, in multiobjective optimization, let $F : \mathbb{R} \to \mathbb{R}^2$,

$$F(x) = \left(x, \sqrt{1 + x^2}\right).$$

Note that F is component-wise convex. In this context, $K = K^* = \mathbb{R}^2_+$, and $y \preccurlyeq_K y'$ means $y_i \leqslant y'_i$, for $i = 1, 2$. Hence, such F is K-convex. If we take $w = (w_1, w_2) > 0$, with $w_1 > w_2$, then the scalarized problem is unbounded. Of course, algorithms for choosing w are quite desirable.

Once we have a very simple example in which the "wrong" choice of w breaks down the method, a very natural question is how does the K-steepest descent method behave in this example. First of all, observe that, in the above example, the set of K-critical points is given by the halfline $(-\infty, 0]$; furthermore, it coincides with the K-optimal set. Let $\{x^k\}$ be the sequence generated by Algorithm 2. If this sequence is finite, then the last iterate is K-critical, and therefore optimal. Suppose that the algorithm does not stop. Observe that, in this example, for any $\tilde{x} \in \mathbb{R}^2$, the set

$$\{x \in \mathbb{R}^2 \mid F(x) \leqslant F(\tilde{x})\}$$

is bounded. As $\{F(x^k)\}$ is K-nonincreasing, it follows that $\{x^k\}$ is bounded. So, it has accumulations points, all of which, by Theorem 4.2, are K-critical (hence K-optimal). Furthermore, if $\bar{x}, \hat{x}$ are accumulation points, then $F(\bar{x}) = F(\hat{x})$. This readily implies (in this particular example) $\hat{x} = \bar{x}$. So $\{x^k\}$ converges to a solution.

It would be desirable to combine some ideas presented in the preceding sections with the scalarization method. Indeed, they are connected. Observe that, for g defined as in (32),

$$\nabla g(x) = JF(x)^t w.$$

So, the steepest descent direction for g at x is $-JF(x)^t w$. Take $\bar{x} \in \mathbb{R}^n$. We claim that for a suitable $w \in K^*$, the steepest descent direction for the scalarized objective function g at $\bar{x}$ coincides with the K-steepest descent direction $v_{\bar{x}}$. To prove this claim, define

$$\widetilde{C} = \mathrm{conv}(C).$$

Obviously, $\widetilde{C}$ is a convex compact set and $\varphi(y) = \sup_{w \in \widetilde{C}} \langle w, y \rangle$. The K-steepest descent direction at $\bar{x}$ is the solution of (17) with $x = \bar{x}$, which may also be written as

$$\min_{v \in \mathbb{R}^n} \max_{w \in \widetilde{C}} \left(\langle w, JF(\bar{x})v \rangle + (1/2)\|v\|^2 \right). \tag{33}$$

The dual of this problem is

$$\max_{w\in\widetilde{C}} \min_{v\in\mathbb{R}^n} (\langle w, JF(\bar{x})v\rangle + (1/2)\|v\|^2). \tag{34}$$

Trivially,

$$\arg\min_{v\in\mathbb{R}^n} \quad \langle w, JF(\bar{x})v\rangle + (1/2)\|v\|^2 = -JF(\bar{x})^t w,$$
$$\min_{v\in\mathbb{R}^n} \quad \langle w, JF(\bar{x})v\rangle + (1/2)\|v\|^2 = (-1/2)\|JF(\bar{x})^t w\|^2.$$

Hence, problem (34) may be simplified to

$$\max_{w\in\widetilde{C}} \quad -(1/2)\|JF(\bar{x})^t w\|^2 \tag{35}$$

or, equivalently,

$$\min_{w\in\widetilde{C}} \quad (1/2)\|JF(\bar{x})^t w\|^2. \tag{36}$$

Since $\widetilde{C}$ is convex and compact, problem (34) has always a solution, say $w_{\bar{x}}$ (which may not be unique) and there is no duality gap. In particular, $(v_{\bar{x}}, w_{\bar{x}})$ is a saddle point of $\langle w, JF(\bar{x})v\rangle + (1/2)\|v\|^2$ in $\mathbb{R}^n \times \widetilde{C}$:

$$\langle w, JF(\bar{x})v_{\bar{x}}\rangle + (1/2)\|v_{\bar{x}}\|^2 \leqslant \langle w_{\bar{x}}, JF(\bar{x})v_{\bar{x}}\rangle + (1/2)\|v_{\bar{x}}\|^2 \leqslant \langle w_{\bar{x}}, JF(\bar{x})v\rangle + (1/2)\|v\|^2$$

for all $(v, w) \in \mathbb{R}^n \times \widetilde{C}$. So,

$$v_{\bar{x}} = -JF(\bar{x})^t w_{\bar{x}} \tag{37}$$

and $w_{\bar{x}} \in \widetilde{C} \subseteq K^*$. Taking $w = w_{\bar{x}}$ in (32) we get $-\nabla g(x) = -JF(x)^t w_{\bar{x}}$, and so $-\nabla g(\bar{x}) = v_{\bar{x}}$, as we claimed.

Since there is no duality gap in (33)–(35), using (37) we get,

$$\alpha_{\bar{x}} = -(1/2)\|v_{\bar{x}}\|^2,$$

where $\alpha_{\bar{x}}$ is the optimal value of problem (17) for $x = \bar{x}$, so

$$\varphi(JF(\bar{x})v_{\bar{x}}) = -\|v_{\bar{x}}\|^2.$$

Now we will prove that approximate solutions of the dual problem (34) (or (35), (36)) yield approximate solutions of (33), i.e., approximate *K*-steepest descent directions at *x*.

Proposition 5.1. *Let* $x \in \mathbb{R}^n$ *be non K-critical and* $\sigma \in (0, 1)$. *There exists* $\delta > 0$ *such that, if* $\tilde{w} \in \widetilde{C}$ *and*

$$(1/2)\|JF(x)^t \tilde{w}\|^2 - (1/2)\|JF(x)^t w_x\|^2 \leqslant \delta,$$

then $\tilde{v} = -JF(x)^t \tilde{w}$ *is a* σ*-approximate K-steepest descent direction at x.*

Proof. Define

$$\varepsilon = -\sigma\alpha_x.$$

As x is non K-critical, $\alpha_x < 0$, and so $\varepsilon > 0$. Since the objective function on (12) and JF are continuous, there exists $\eta > 0$ such that

$$\|v - v_x\| \leqslant \eta \;\;\Rightarrow\;\; \varphi(JF(x)v) + (1/2)\|v\|^2 \leqslant \varphi(JF(x)v_x) + (1/2)\|v_x\|^2 + \varepsilon.$$

So, $\|v - v_x\| \leqslant \eta$ implies that v is a σ-approximate K-steepest descent direction at x.

Optimality of w_x for (36), convexity of $\widetilde{C}$ and the hypothesis $\tilde{w} \in \widetilde{C}$ imply

$$\langle JF(x)^t(\tilde{w} - w_x), JF(x)^t w_x\rangle \geqslant 0.$$

Hence, using the equalities $\tilde{v} = -JF(x)^t\tilde{w}$, $v_x = -JF(x)^t w_x$,

$$\begin{aligned}\|JF(x)^t\tilde{w}\|^2 &\geqslant \|JF(x)^t w_x\|^2 + \|JF(x)^t(\tilde{w} - w_x)\|^2\\ &= \|JF(x)^t w_x\|^2 + \|\tilde{v} - v_x\|^2.\end{aligned}$$

Therefore,

$$\|\tilde{v} - v_x\| \leqslant \sqrt{\|JF(x)^t\tilde{w}\|^2 - \|JF(x)^t w_x\|^2}.$$

So, it is enough to take $\delta = \eta^2/2 > 0$. $\square$

Let $w \in \widetilde{C}$ be an approximate solution of (35) (or (36)). A question of practical relevance is whether $v = -JF(x)^t\tilde{w}$ is a σ-approximate K-steepest descent direction at x. In our next proposition we give a sufficient condition for σ-approximation.

Proposition 5.2. *Take $\sigma \in [0, 1)$, $w \in \widetilde{C}$ and define $v = -JF(x)^t w$. If*

$$\varphi(JF(x)v) \leqslant -(1 - \sigma/2)\|v\|^2,$$

then v is a σ-approximate K-steepest descent direction at x.

Proof. We already know that (17) (or (33)) and (35) are a primal-dual pair of problems. Since w is dual feasible,

$$-(1/2)\|JF(x)^t w\|^2 \leqslant \alpha_x.$$

Therefore, making the substitution $v = -JF(x)^t w$ we get

$$(1 - \sigma)(-1/2)\|v\|^2 \leqslant (1 - \sigma)\alpha_x. \tag{38}$$

Hence, if

$$\varphi(JF(x)v) \leqslant -(1 - \sigma/2)\|v\|^2,$$

from (38) it follows that $v = -JF(x)^t w$ is a σ-approximate K-steepest descent direction. $\square$

In the following section, we will show a theoretical advantage of using descent directions as discussed in Propositions 5.1, 5.2. From now on, we will say that v is *scalarization compatible*, or *s-compatible* if there exist some $w \in \widetilde{C}$ such that

$$v = -JF(x)^t w = \nabla_x\langle w, F(x)\rangle.$$

Note that v_x, the *exact* K-steepest descent direction at x, is always s-compatible.

6. Convergence analysis: the K-convex case

Throughout this section we will assume that F is K-convex, i.e.,

$$F(\lambda x + (1-\lambda)x') \preccurlyeq_K \lambda F(x) + (1-\lambda)F(x')$$

for all $x, x' \in \mathbb{R}^n$ and all $\lambda \in [0, 1]$. Under this additional assumption we have the following extension of the classical gradient inequality to the vector case:

$$F(x) + JF(x)(x' - x) \preccurlyeq_K F(x'),$$

for any $x, x' \in \mathbb{R}^n$ (see [19, Lemma 5.2]).

As in the general case, here also optimality implies K-criticality. A point $x^* \in \mathbb{R}^n$ is a *weak unconstrained K-minimizer* (or *weak K-optimum*) of F, or weak Pareto minimal element for F (see [19]), if there is no $x \in \mathbb{R}^n$ with $F(x) \prec_K F(x^*)$. A well known fact is that, under the K-convexity assumption on the objective function F, K-criticality and *weak* optimality are equivalent conditions.

If the algorithm has finite termination, the last iterate is K-critical and hence a weak unconstrained K-minimizer of F. We will study the case in which the algorithm does not have finite termination and therefore produces infinite sequences $\{x^k\}$, $\{v^k\}$ and $\{t_k\}$. Let us now establish the additional assumptions under which we will prove full convergence of $\{x^k\}$ to a K-critical point or, in view of the above discussion, to a weak unconstrained K-minimizer of F.

A1. Every K-decreasing sequence in the image of F

$$\{y^k\} \subseteq \{F(x) \mid x \in \mathbb{R}^n\}$$

is K-bounded below by a point in the image of F.

A2. All v^k's are scalarization compatible, i.e., there exists a sequence $\{w^k\}$ in $\widetilde{C}$ such that,

$$v^k = -JF(x^k)^t w^k, \quad k = 0, 1, \ldots$$

Some comments concerning the generality/restrictiveness of these assumptions are in order. Regarding Assumption A1, in the case of classical unconstrained (convex) optimization, this condition is equivalent to existence of solutions of the optimization problem. This assumption, known as *K-completeness*, is standard for ensuring existence of efficient points for vector optimization problems (see [19, Section 3]). Assumption A2 deals with the implementation of the algorithm rather than with F. This assumption holds if each v^k is the *exact* K-steepest descent direction at x^k (see (37)). Of course, Assumption A2 applies to the sequence of directions $\{v^k\}$ prescribed by the algorithm, i.e., we are assuming that v^k is an s-compatible σ-approximate K-steepest descent direction for all k.

We will need the following technical lemma in order to prove that the K-steepest descent method is convergent.

Lemma 6.1. *Suppose that F is K-convex and that v^k is scalarization compatible (at x^k). If $F(\hat{x}) \preccurlyeq_K F(x^k)$ then*

$$\|\hat{x} - x^{k+1}\|^2 \leqslant \|\hat{x} - x^k\|^2 + \|x^{k+1} - x^k\|^2.$$

Proof. By assumption, there exists some $w^k \in \widetilde{C}$ such that

$$v^k = -JF(x^k)^t w^k.$$

Using the K-convexity of F we have $F(x^k) + JF(x^k)(\hat{x} - x^k) \preccurlyeq_K F(\hat{x})$. Since $F(\hat{x}) \preccurlyeq_K F(x^k)$, we get

$$JF(x^k)(\hat{x} - x^k) \preccurlyeq_K 0.$$

Taking into account that $w^k \in K^*$ and using the above results we get

$$-(v^k)^t(\hat{x} - x^k) = (w^k)^t JF(x^k)(\hat{x} - x^k) \leqslant 0.$$

Recall that $x^{k+1} = x^k + t_k v^k$, with $t_k > 0$. Therefore

$$(x^k - x^{k+1})^t(\hat{x} - x^k) \leqslant 0,$$

which implies the desired inequality, because

$$\|\hat{x} - x^{k+1}\|^2 = \|\hat{x} - x^k\|^2 + \|x^k - x^{k+1}\|^2 + 2(x^k - x^{k+1})^t(\hat{x} - x^k). \qquad \square$$

Before stating our convergence result, we recall that a sequence $\{y^k\} \subset \mathbb{R}^m$ is *quasi-Fejér convergent* [10,11] to a set $U \subset \mathbb{R}^m$ if for every $u \in U$ there exists a sequence $\{\varepsilon_k\} \subset \mathbb{R}$, $\varepsilon_k \geqslant 0$ such that

$$\|y^{k+1} - u\|^2 \leqslant \|y^k - u\|^2 + \varepsilon_k \quad \text{for all } k = 1, 2, \ldots,$$

with

$$\sum_{k=1}^{\infty} \varepsilon_k < \infty.$$

We will also need the following result concerning quasi-Fejér convergent sequences, whose proof can be found in [4,18].

Theorem 6.2. *If the sequence $\{y^k\}$ is quasi-Fejér convergent to a nonempty set $U \subset \mathbb{R}^m$, then $\{y^k\}$ is bounded. If furthermore a cluster point y of $\{y^k\}$ belongs to U, then* $\lim_{k\to\infty} y^k = y$.

In [4], it is proved that the steepest descent method for smooth (scalar) convex minimization, with stepsize obtained using backtracking and Armijo rule, is globally convergent to a solution (under the sole assumption of existence of optima). We will extend those results to the K-steepest descent method, using the same techniques as in [4].

Theorem 6.3. *Suppose that F is K-convex and that Assumptions* A1, A2 *hold. Then $\{x^k\}$ converges to a K-critical point x^*.*

Proof. First of all, observe that all results of Section 4 are still valid under the additional assumptions of this theorem. In particular, $\{F(x^k)\}$ is a K-decreasing sequence, so, using Assumptions A1, there exists an $\bar{x} \in \mathbb{R}^n$ such that,

$$F(\bar{x}) \preccurlyeq_K F(x^k) \quad \forall k \in \mathbb{N}. \tag{39}$$

Now observe that $0 < t_k \leqslant 1$ for all k. Hence

$$t_k \| v(x^k) \|^2 = \frac{1}{t_k} \| x^{k+1} - x^k \|^2 \geqslant \| x^{k+1} - x^k \|^2. \tag{40}$$

Therefore, from (39), (40) and Lemma 4.1 it follows that

$$\sum_{k=1}^{\infty} \| x^k - x^{k+1} \|^2 < \infty.$$

Define

$$L := \{ x \in \mathbb{R}^n \mid F(x) \preccurlyeq_K F(x^k) \quad \forall k \in \mathbb{N} \}.$$

Note that $\bar{x} \in L$, so L is nonempty. Using Assumption A2 and Lemma 6.1, we conclude that for any $x \in L$ (and $k \in \mathbb{N}$),

$$\| x - x^{k+1} \|^2 \leqslant \| x - x^k \|^2 + \| x^k - x^{k+1} \|^2.$$

Since $\sum_{k=1}^{\infty} \| x^k - x^{k+1} \|^2 < \infty$, we conclude that the sequence $\{x^k\}$ is quasi-Fejér convergent to the set L. As L is nonempty, from Theorem 6.2 it follows that $\{x^k\}$ has accumulation points. Let x^* be one of them. By Proposition 2.2, $x^* \in L$. Then, once again by virtue of Theorem 6.2, it follows that the whole sequence $\{x^k\}$ converges to x^*. We finish the proof by observing that Theorem 4.2 guarantees that x^* is K-critical. □

7. Vector optimization and the abstract equilibrium problem

In this section, we discuss the connection between the problem of seeking a weak constrained K-minimizer and the Abstract Equilibrium problem. We will see that by means of the function φ, defined in (12), the first problem can be viewed as a particular case of the second one.

The *Weak Constrained K-Minimization problem* [19] is defined in the following way:

Given closed convex pointed cones with nonempty interior $K \subset \mathbb{R}^m$, $K_i \subset \mathbb{R}^{m_i}$, $i = 1, 2, \ldots, r$, the corresponding induced orders: $x \prec_K y$ if $y - x \in \operatorname{int}(K)$, $u \preccurlyeq_{K_i} v$ if $v - u \in K_i$, and

$F : \mathbb{R}^n \to \mathbb{R}^m$, K-convex and continuously differentiable,

$G_i : \mathbb{R}^n \to \mathbb{R}^{m_i}$, K_i-convex, for all $i = 1, 2, \ldots, r$,

find $x^* \in M = \{ x \in \mathbb{R}^n \mid G_i(x) \preccurlyeq_{K_i} 0$, for $1 \leqslant i \leqslant r \}$ such that there does not exist any other $x \in M$ with, $F(x) \prec_K F(x^*)$.

The *Abstract Equilibrium problem* [3] can be stated in the following way: Given M, a nonempty closed convex set in a Hausdorff topological vector space and

$$f : M \times M \to \mathbb{R}$$

such that

$f(\cdot, y)$ is upper semicontinuous for all $y \in M$,

$f(x, \cdot)$ is convex and lower semicontinuous for all $x \in M$,

$f(x, x) = 0$ for all $x \in M$,

find $\bar{x} \in M$ with the property that

$$f(\bar{x}, y) \geqslant 0 \quad \text{for all } y \in M.$$

Taking

$$M = \{x \in \mathbb{R}^n \mid G_i(x) \preccurlyeq_{K_i} 0 \quad \text{for } 1 \leqslant i \leqslant r\}$$

as in the weak constrained K-minimization problem, and

$$f(x, y) := \varphi(F(y) - F(x)),$$

where φ was defined in (12), we have that the solution set of this equilibrium problem is exactly the same as the solution set of the weak constrained K-minimization problem. Observe that the weak unconstrained K-minimization problem is a particular case of the constrained one and, therefore, can also be solved by means of methods for solving the equilibrium problem (and conversely). So we can find K-critical points for problem (1) via the equilibrium problem formulation.

8. Final remarks

In this work we proposed for vector unconstrained minimization an extension of the standard steepest descent method. We showed that all cluster points of the sequences produced by the method satisfy a certain first-order condition for K-optimality, known as K-criticality. Under K-convexity of the objective function and assuming a very reasonable condition on the objective function, we proved that we have full convergence of the method, when performed, for example, with the exact K-steepest descent direction at each iteration. In this situation, no matter how bad is our initializing point, the method will converge to a K-critical point, or, in other words, to a weak unconstrained K-minimizer. As in the one-dimensional case, using the notion of quasi-Fejér convergence, we could prove convergence of the method without assuming a Lipschitz condition on the objective's Jacobian and without bounded level sets requirement. Incidentally, we showed that every weak constrained K-minimization problem can be viewed as a particular equilibrium problem.

It is worth mentioning that the compact set $C \subset K^*$ used throughout the whole work is essential for defining the K-steepest descent direction (or approximations of it). So, somehow, C plays the role of a sort of "gauge". Perhaps, it would have been better to define the "gauged" K-steepest descent direction, or the K-steepest descent direction modulus C, instead of simply the K-steepest descent direction. This would make the notation more clear but certainly heavier.

Regarding the implementation of the method, we mention that when K (therefore K^*) is finitely generated, the K-steepest descent direction can be easily computed. In the general case, approximate K-steepest descent direction can be computed solving approximately (36) and using Proposition 5.1 or Proposition 5.2. Other possibility is to solve (17) using a bundle method technique [17].

We expect that, in spite of the possible drawbacks of the method, as in the single-valued case, it will furnish a prototype for more sophisticated and efficient algorithms for solving vector optimization problems. So we think that a full understanding of the structure and convergence behavior of the K-steepest descent method is indeed relevant.

The extension to vector optimization of more efficient algorithms, as, for instance, Newton's and quasi-Newton methods, and its convergence analysis are left as open problems for future research.

References

[1] C.D. Aliprantis, M. Florenzano, V.F. Martins da Rocha, R. Tourky, Equilibrium analysis in financial markets with countably many securities, J. Math. Econom. 40 (2004) 683–699.

[2] C.D. Aliprantis, M. Florenzano, R. Tourky, General equilibrium analysis in ordered topological vector spaces, J. Math. Econom. 40 (3–4) (2004) 247–269.

[3] E. Blum, W. Oettli, From optimization and variational inequalities to equilibrium problems, Math. Student 63 (1–4) (1994) 123–145.

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

[5] E. Carrizosa, E. Conde, M. Muñoz Márquez, J. Puerto, Planar point-objective location problems with nonconvex constraints: a geometrical construction, J. Global Optim. 6 (1) (1995) 77–86.

[6] E. Carrizosa, J.B.G. Frenk, Dominating sets for convex functions with some applications, J. Optim. Theory Appl. 96 (2) (1998) 281–295.

[7] I. Das, J.E. Dennis, A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multi-criteria optimization problems, Structural Optim. 14 (1) (1997) 63–69.

[8] I. Das, J.E. Dennis, Normal-boundary intersection: a new method for generating the Pareto surface in nonlinear multicriteria optimization problems, SIAM J. Optim. 8 (3) (1998) 631–657 (electronic).

[9] J.E. Dennis Jr., R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996 (Corrected reprint of the 1983 original).

[10] Ju.M. Ermol'ev, On the method of generalized stochastic gradients and stochastic quasi-Féjer sequences, Kibernetika (Kiev) 1969 (2) (1969) 73–83.

[11] Ju. M. Ermol'ev, A.D. Tuniev, Random Féjer and quasi-Féjer sequences, in: Theory of Optimal Solutions (Proc. Sem., Kiev, 1968), No. 2 (Russian), Akad. Nauk Ukrain. SSR, Kiev, 1968, pp. 76–83.

[12] H. Eschenauer, J. Koski, A. Osyczka, Multicriteria Design Optimization, Springer, Berlin, 1990.

[13] G.W. Evans, An overview of techniques for solving multiobjective mathematical programs, Management Sci. 30 (11) (1984) 1268–1282.

[14] J. Fliege, B.F. Svaiter, Steepest descent methods for multicriteria optimization, Math. Methods Oper. Res. 51 (3) (2000) 479–494.

[15] A.M. Geoffrion, Proper efficiency and the theory of vector maximization, J. Math. Anal. Appl. 22 (1968) 618–630.

[16] Y.Y. Haimes, L.S. Lasdon, D.A. Wismer, On a bicriterion formulation of the problems of integrated system identification and system optimization, IEEE Trans. Systems Man Cybernet. SMC-1 (1971) 296–297.

[17] J.B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms, II, Fundamental Principles of Mathematical Science, Springer, Berlin, 1993.

[18] A.N. Iusem, B.F. Svaiter, M. Teboulle, Entropy-like proximal methods in convex programming, Math. Oper. Res. 19 (4) (1994) 790–814.

[19] T.D. Luc, Theory of Vector Optimization, Lecture Notes in Economics and Mathematical Systems, vol. 319, Springer, Berlin, 1989.

[20] K. Miettinen, M.M. Mäkelä, Interactive bundle-based method for nondifferentiable multiobjective optimization: Nimbus, Optimization 34 (1995) 231–246.

[21] K.M. Miettinen, Nonlinear Multiobjective Optimization, Kluwer Academic Publishers, Norwell, 1999.

[22] R.T. Rockafellar, Convex Analysis, Princeton University Press, New Jersey, 1970.

[23] C. Tammer, K. Winkler, A new scalarization approach and applications in multicriteria, J. Nonlinear Convex Anal. 4 (3) (2003) 365–380.

[24] D.J. White, Epsilon-dominating solutions in mean-variance portfolio analysis, European J. Oper. Res. 105 (1998) 457–466.