A SURVEY ON MULTIOBJECTIVE DESCENT METHODS

Ellen H. Fukuda¹ and Luis Mauricio Graña Drummond²*

Received September 28, 2013 / Accepted January 2, 2014

ABSTRACT. We present a rigorous and comprehensive survey on extensions to the multicriteria setting of three well-known scalar optimization algorithms. Multiobjective versions of the steepest descent, the projected gradient and the Newton methods are analyzed in detail. At each iteration, the search directions of these methods are computed by solving real-valued optimization problems and, in order to guarantee an adequate objective value decrease, Armijo-like rules are implemented by means of a backtracking procedure. Under standard assumptions, convergence to Pareto (weak Pareto) optima is established. For the Newton method, superlinear convergence is proved and, assuming Lipschitz continuity of the objectives second derivatives, it is shown that the rate is quadratic.

Keywords: multiobjective optimization, Newton method, nonlinear optimization, projected gradient method, steepest descent method.

1 INTRODUCTION

In multiobjective (or multicriteria) optimization, finitely many objective functions have to be minimized simultaneously. Hardly ever a single point will minimize all of them at once, so we need another notion of optimality. Here, we will use the concepts of Pareto and weak Pareto optimality. A point is called Pareto optimal or efficient, if there does not exist a different point with the same or smaller objective function values, such that there is a strict decrease in at least one objective function value. A point is called weakly Pareto optimal or weakly efficient if there does not exist a different point with strict decrease in all its objective values. Applications for this type of problem can be found in engineering [14] (especially truss optimization [12, 32], design [20, 31], space exploration [39, 41]), statistics [10], management science [15, 40, 42], environmental analysis [34, 16], etc.

One of the main solution strategies for multiobjective optimization problems is the scalarization approach [23, 24, 43]: we obtain the optimal solutions of the vector-valued problem by solving...
one or several parametrized single-objective (i.e., scalar-valued) optimization problems. Among
the scalarization techniques, we have the so-called weighting method, where one minimizes a
nonnegative linear combination of the objective functions \([26, 29, 30, 37]\). The parameters are
not known in advance and the modeler or decision-maker has to choose them. For some prob-
lems, this choice can be problematic, as shown in the example provided in \([18, \text{Section 7}]\), where
almost all choices of the parameters lead to unbounded scalar problems. Only recently, adap-
tive scalarization techniques, where scalarization parameters are chosen automatically during the
course of the algorithm such that a certain quality of approximation is maintained, have been
proposed \([13, 17]\). Still other techniques, working only in the bicriteria case, can be viewed as
choosing a fixed grid in a particular parameter space \([11, 12]\).

Multicriteria optimization algorithms that do not scalarize have recently been developed (see,
e.g., \([5, 6]\) for an overview on the subject). Some of these techniques are extensions of scalar
optimization algorithms (notably the steepest descent algorithm \([19]\) with at most linear con-
vergence), while others borrow heavily from ideas developed in heuristic optimization \([33, 38]\). For
the latter, no convergence proofs are known, and empirical results show that convergence gen-
erally is, as in the scalar case, quite slow \([44]\). Other parameter-free multicriteria optimization
techniques use an ordering of the different criteria, i.e., an ordering of importance of the compo-
nents of the objective function vector. In this case, the ordering has to be prespecified. Moreover,
the corresponding optimization process is usually augmented by an interactive procedure \([36]\).

Following the research line opened in 2001 with \([19]\), other classical (scalar) optimization pro-
cedures where extended in recent years, not just for the multicriteria setting, but also for vector
optimization problems, i.e., when other underlying ordering cones are used instead of the non-
negative orthant. In \([28]\), a steepest descent method is proposed, while in \([1, 21, 22, 25]\), several
versions of the projected gradient method are studied. Proximal point type methods are analyzed
in \([4]\). Trust region strategies for multicriteria were developed in \([8, 9]\). Even a steepest descent
method for multicriteria optimization on Riemannian manifolds was developed in \([2]\). In 2009,
Fliege et al. \([18]\) came up with a multiobjective version of the Newton method, while later on,
Gra\~{n}a Drummond et al. \([27]\) proposed a Newton method for vector optimization.

These extensions share an essential feature: they are all descent methods, i.e., the vector objec-
tive value decreases at each iteration in the partial order induced by the underlying cones. Under reasonable assumptions, full convergence of sequences produced by these algorithms is estab-
lished in all those works. As expected, for the vector-valued Newton methods, convergence is
local and fast, while for the others it is global (and not so fast).

In this survey we study just three of these procedures: the steepest descent \([19, 28]\), the projected
gradient \([21, 22, 25]\) and the Newton \([18, 27]\) methods. For the sake of simplicity, we just analyze
their multiobjective versions and, in order to emphasize the underlying ideas and avoid techni-
calities, we do not present the results in their maximum degree of generality. For instance, we
assume that the objectives are continuously differentiable in the whole space, while, sometimes,
we just need local differentiability, etc.
The outline of this paper is as follows. In Section 2 we establish the notation, as well as the basic concepts, and define the (smooth) unconstrained multiobjective problem. We also state and prove a simple but important result on descent directions. In Section 3, we study the multiobjective steepest descent method. First, we introduce the search direction and analyze its properties. Then we describe the algorithm and, finally, we perform the convergence analysis of the method. Basically, we establish that, under convexity of the objectives and another natural assumption, any sequence produced by the method is globally convergent to a weak Pareto optimal point. In Section 4 we define the smooth constrained multicriteria problem, as well as the multiobjective projected gradient algorithm and we present its convergence analysis, which is quite similar to the previous one. In Section 5, for twice continuously differentiable strongly convex objectives, we define the multiobjective Newton method and, assuming similar conditions as those of the previous cases, we establish its convergence to Pareto optimal points with superlinear rate. Under the additional hypothesis of Lipschitz continuity of the objectives’ second derivatives, we prove quadratic convergence. In Section 6 we make some comments on how to obtain (strong) Pareto optima with the first two algorithms and on how to approach the whole efficient frontier with the three methods.

2 PRELIMINARIES

All over this work, \( \langle \cdot, \cdot \rangle \) will stand for the usual canonical inner product in \( \mathbb{R}^p \) (here, we will just deal with \( p = n \) or \( p = m \)), i.e., \( \langle u, v \rangle = \sum_{i=1}^{p} u_i v_i = u^\top v \). Likewise, \( \| \cdot \| \) will always be the Euclidean norm, i.e., \( \|u\| = \sqrt{\langle u, u \rangle} \). For an \( m \times n \) real matrix \( A \), \( \| A \| \) will denote its operator (Euclidean) norm. As usual, \( \mathbb{R}^+ \) and \( \mathbb{R}^{++} \) designate the sets of nonnegative and positive real numbers, respectively.

Let \( \mathbb{R}_+^m := \mathbb{R}_+ \times \cdots \times \mathbb{R}_+ \) and \( \mathbb{R}^{++}_m := \mathbb{R}^{++} \times \cdots \times \mathbb{R}^{++} \) be the nonnegative orthant or Paretian cone and the positive orthant of \( \mathbb{R}^m \), respectively. Consider the partial order induced by \( \mathbb{R}^+_m \): for \( u, v \in \mathbb{R}^m \), \( u \leq v \) (alternatively, \( v \geq u \)) if \( v - u \in \mathbb{R}^+_m \), as well as the following stronger relation: \( u < v \) (alternatively, \( v > u \)) if \( v - u \in \mathbb{R}^{++}_m \). In other words, \( u \leq v \) stands for \( u_i \leq v_i \) for all \( i = 1, \ldots, m \), and \( u < v \) should also be understood as a (strict) componentwise inequality.

Let 
\[
F: \mathbb{R}^n \rightarrow \mathbb{R}^m
\]
be a continuously differentiable mapping. Our problem consists of finding Pareto or weak Pareto points for \( F \) in \( \mathbb{R}^n \) and we note it by

\begin{equation}
\begin{aligned}
\text{minimize} \quad & F(x) \\
\text{subject to} \quad & x \in \mathbb{R}^n.
\end{aligned}
\end{equation}

Recall that \( x^* \in \mathbb{R}^n \) is a Pareto optimal (or efficient) point for \( F \), if

there is no \( x \in \mathbb{R}^n \) such that \( F(x) \leq F(x^*) \) and \( F(x) \neq F(x^*) \),
that is to say, there does not exist \( x \in \mathbb{R}^n \) such that \( F_i(x) \leq F_i(x^*) \) for all \( i = 1, \ldots, m \) and \( F_h(x) < F_h(x^*) \) for at least one \( h \). A point \( x^* \in \mathbb{R}^n \) is a weakly Pareto optimal (or weakly efficient) point for \( F \), if

there is no \( x \in \mathbb{R}^n \) such that \( F(x) < F(x^*) \),

i.e., there does not exist \( x \in \mathbb{R}^n \) such that \( F_i(x) < F_i(x^*) \) for all \( i = 1, \ldots, m \). Clearly, if \( x^* \in \mathbb{R}^n \) is a Pareto optimal point, then \( x^* \) is a weak Pareto point, and the converse is not always true.

As we will see in our first lemma, a necessary condition for optimality of a point \( \tilde{x} \in C \) is stationarity (or criticality):

\[
JF(\tilde{x})(\mathbb{R}^n) \cap [-\mathbb{R}^n_{++}] = \emptyset,
\]

where \( JF(\tilde{x}) \) stands for the Jacobian matrix of \( F \) at \( \tilde{x} \) (the \( m \times n \) matrix with entries \( (JF(\tilde{x}))_{i,j} = \partial F_i(\tilde{x})/\partial x_j \)). \( JF(\tilde{x})(\mathbb{R}^n) := \{ JF(\tilde{x})v : v \in \mathbb{R}^n \} \) and \( -\mathbb{R}^n_{++} := \{ -u : u \in \mathbb{R}^m_{++} \} \). So \( \tilde{x} \in \mathbb{R}^n \) is stationary for \( F \) if, and only if, for all \( v \in \mathbb{R}^n \) we have \( JF(\tilde{x})v \neq 0 \), that is to say for all \( v \) there exists \( j = j(v) \) such that

\[
(\sum_{i=1}^{m} \frac{\partial F_i}{\partial x_j} v_J) \geq 0,
\]

and this condition implies

\[
\max_{i=1,\ldots,m} \nabla F_i(\tilde{x})^T v \geq 0 \quad \text{for all } v \in \mathbb{R}^n.
\]

Note that for \( m = 1 \), we retrieve the classical stationarity condition for unconstrained scalar-valued optimization: \( \nabla F(\tilde{x}) = 0 \).

By definition, a point \( x \in \mathbb{R}^n \) is nonstationary if there exists a direction \( v \in \mathbb{R}^n \) such that \( JF(x)v < 0 \). The next result shows not only that such \( v \) is a \( \mathbb{R}^n_{++} \)-descent direction for \( F \) at \( x \) (there exists \( \varepsilon > 0 \) such that \( F(x+tv) < F(x) \) for all \( t \in (0, \varepsilon) \)), but it also estimates the \( \mathbb{R}^n_{++} \)-function decrease. As we will see later on, this result allows us to implement, by a backtracking procedure, an Armijo-like rule on all the multiobjective (descent) methods we will study here.

**Lemma 2.1.** Let \( F : \mathbb{R}^n \to \mathbb{R}^m \) be continuously differentiable, \( v \in \mathbb{R}^n \) such that \( JF(x)v < 0 \) and \( \sigma \in (0, 1) \). Then there exists some \( \varepsilon > 0 \) (which may depend on \( x \), \( v \) and \( \sigma \)) such that

\[
F(x+tv) < F(x) + \sigma t JF(x)v \quad \text{for any } t \in (0, \varepsilon].
\]

In particular, \( v \) is a descent direction for \( F \) at \( x \).

**Proof.** Since \( F \) is differentiable, \( \nabla F_i(x)^T v < 0 \) and \( \sigma \in (0, 1) \), we have

\[
\lim_{t \to 0} \frac{F_i(x+tv) - F_i(x)}{t} = \nabla F_i(x)^T v < \sigma \nabla F_i(x)^T v \quad \text{for all } i = 1, \ldots, m.
\]

So there exists \( \varepsilon > 0 \) such that

\[
F_i(x+tv) - F_i(x) < \sigma t \nabla F_i(x)^T v \quad \text{for all } t \in (0, \varepsilon] \quad \text{and all } i = 1, \ldots, m.
\]
3 THE MULTIOBJECTIVE STEEPEST DESCENT METHOD

In this section we propose an extension of the classical steepest descent method. First we state and prove some results that will allow us to define the algorithm. Then we present its convergence analysis and, finally, we briefly expose a numerically robust version of the method which admits relative errors on computing the search directions at each iteration.

3.1 The search direction for the multiobjective steepest descent method

For a given point \( x \in \mathbb{R}^n \), define the function \( \varphi_x : \mathbb{R}^n \rightarrow \mathbb{R} \) by
\[
\varphi_x(v) := \max_{i=1,...,m} \nabla F_i(x)^\top v. \tag{4}
\]
As being the maximum of linear functions, \( \varphi_x \) is convex and positively homogeneous (\( \varphi_x(\lambda v) = \lambda \varphi_x(v) \) for all \( \lambda > 0 \) and all \( v \in \mathbb{R}^n \)). Moreover, since \( u \mapsto \max_{i=1,...,m} u_i \) is Lipschitz continuous with constant 1, we also have
\[
|\varphi_x(v + w) - \varphi_x(v)| \leq \|JF(x)v - JF(y)w\|, \tag{5}
\]
for all \( v, w \in \mathbb{R}^n \). Consider now the unconstrained scalar-valued minimization problem
\[
\min_v \varphi_x(v) + \frac{1}{2}\|v\|^2. \tag{6}
\]
Since the objective function is proper, closed and strongly convex, this problem has always a (unique) optimal solution \( v(x) \), which we call steepest descent direction, and is therefore given by
\[
v(x) := \arg\min_{v \in \mathbb{R}^n} \varphi_x(v) + \frac{1}{2}\|v\|^2. \tag{7}
\]
Let us call \( \theta(x) \) the optimal value of (6), that is to say,
\[
\theta(x) := \varphi_x(v(x)) + \frac{1}{2}\|v(x)\|^2. \tag{8}
\]
If \( m = 1 \), we fall back to scalar minimization and \( \varphi_x(v) = \nabla F(x)^\top v \), so \( v(x) = -\nabla F(x) \) and we retrieve the classical steepest descent direction at \( x \in \mathbb{R}^n \).

Note that, in order to get rid of the possible nondifferentiability of (6), we can reformulate the problem as
\[
\begin{align*}
\text{minimize} & \quad \tau \\
\text{subject to} & \quad \nabla F_i(x)^\top v + \frac{1}{2}\|v\|^2 - \tau \leq 0, \quad i = 1, \ldots, m,
\end{align*}
\]
a convex problem with variables \( (\tau, v) \in \mathbb{R} \times \mathbb{R}^n \) for which there are efficient numerical solvers.

Let us now give a characterization of stationarity in terms of \( \theta(\cdot) \) and \( v(\cdot) \) and study some features of these functions.
Proposition 3.1. Let \( v: \mathbb{R}^n \to \mathbb{R}^n \) and \( \theta: \mathbb{R}^n \to \mathbb{R} \) given by (7) and (8), respectively. Then, the following statements hold.

1. \( \theta(x) \leq 0 \) for all \( x \in \mathbb{R}^n \).
2. The function \( \theta(\cdot) \) is continuous.
3. The following conditions are equivalent:
   (a) The point \( x \) is nonstationary.
   (b) \( \theta(x) < 0 \).
   (c) \( v(x) \neq 0 \).

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

Proof. 1. From (7) and (8), for any \( x \in \mathbb{R}^n \), we have \( \theta(x) \leq \phi_x(0) + \frac{1}{2} \|0\|^2 = 0 \).

2. Let us prove the continuity of \( \theta(\cdot) \) in an arbitrary but fixed \( x_0 \in \mathbb{R}^n \). First, let us show that the function \( x \mapsto v(x) \) is bounded on compact subsets of \( \mathbb{R}^n \). Let \( C \subset \mathbb{R}^n \) be compact. Note that for all \( x \in C \) and all \( i = 1, \ldots, m \), we have

\[
-\|\nabla F_i(x)\|_2 \|v(x)\|_2 + \frac{1}{2} \|v(x)\|^2_2 \leq \nabla F_i(x)^T v(x) + \frac{1}{2} \|v(x)\|^2_2 \leq \phi_x(v(x)) + \frac{1}{2} \|v(x)\|^2_2 = \theta(x) \leq 0,
\]

where we used the Cauchy-Schwarz inequality, the definitions of \( \phi_x \) and \( \theta \), as well as item 1. Therefore, since \( F \) is continuously differentiable, we conclude that there exists \( \kappa = \kappa(C) > 0 \) such that \( \|v(x)\| \leq \kappa \) for all \( x \in C \). In particular, there exists \( \tilde{\kappa} > 0 \) such that

\[
\|v(x)\| \leq \tilde{\kappa} \quad \text{for all} \quad x \in \tilde{B}(x^o, 1),
\]

where \( \tilde{B}(x^o, 1) \) is the closed ball with center at \( x^o \) and radius 1.

Now, for \( x \in \tilde{B}(x^o, 1) \) and \( i \in \{1, \ldots, m\} \), define

\[
\phi_{x,i}: \mathbb{R}^n \to \mathbb{R}
\]

by

\[
z \mapsto \nabla F_i(z)^T v(x) + \frac{1}{2} \|v(x)\|^2_2.
\]

From (9), the family \( \{\phi_{x,i}\}_{i \in \mathbb{B}(x^o, 1) \times \{1, \ldots, m\}} \) is equicontinuous. Consider now the family \( \{\Phi_u\}_{x \in \tilde{B}(x^o, 1)} \), where \( \Phi_u = \max_{i = 1, \ldots, m} \phi_{x,i} \). Since \( u \mapsto \max_i u_i \) is Lipschitz continuous with

Pesquisa Operacional, Vol. 34(3), 2014
constant 1, \( \{ \Phi_i \}_{i \in B(\varepsilon, 1)} \) is also an equicontinuous family of functions. Then for any \( \varepsilon > 0 \) and all \( x \in \bar{B}(x^0, 1) \), there exists \( 0 < \delta < 1 \) such that
\[
\|z - \bar{z}\| < \delta \quad \Rightarrow \quad |\Phi_i(z) - \Phi_i(\bar{z})| < \varepsilon.
\]
Hence for \( \|z - x^0\| < \delta \), from the definitions of functions \( \theta, \tau, \Phi_{e^0} \) and the fact that \( \Phi_{e^0}(x^0) = \theta(x^0) \), we have
\[
\theta(z) \leq \max_{i=1, \ldots, n} \|v_i(x) + 1/2 \|v(x)\|^2 \leq \Phi_{e^0}(x^0) + |\Phi_{e^0}(z) - \Phi_{e^0}(x^0)| < \theta(x^0) + \varepsilon,
\]
so \( \theta(z) - \theta(x^0) < \varepsilon \). Interchanging the roles of \( z \) and \( x^0 \), we conclude that, for \( \|z - x^0\| < \delta \),
\[
|\theta(z) - \theta(x^0)| < \varepsilon.
\]

3. (a) \( \Rightarrow \) (b). Suppose that \( x \) is nonstationary. Then, there exists \( v \in \mathbb{R}^n \) such that \( \varphi_i(v) = \max \|F(x)^\top v\| < 0 \). Therefore, from the optimality of \( v(x) \) and the fact that \( \varphi_i \) is positively homogeneous,
\[
\theta(x) \leq \varphi_i(v(x)) + 1/2 \|v(x)\|^2 = \tau \left( \varphi_i(v(x)) + 1/2 \|v\|^2 \right)
\]
for all \( \tau > 0 \). Taking \( 0 < \tau < -2\varphi_i(v)/\|v\|^2 \), we see that \( \theta(x) < 0 \).

(b) \( \Rightarrow \) (c). Since, by item 1, \( \theta(x) \leq 0 \), it is easy to notice that \( v(x) = 0 \) implies \( \theta(x) = 0 \).

(c) \( \Rightarrow \) (a). If \( v(x) \neq 0 \), then \( \max_i \|F_i(x)^\top v(x)\| = \varphi_i(v(x)) < \theta(x) \leq 0 \), so \( JF(x)v(x) \in JF(x)(\mathbb{R}^n) \cap [-\mathbb{R}^{+\infty}, 0] \) and \( x \) is nonstationary. \( \square \)

Note that for a nonstationary point \( x \), using the above characterization, we have \( \theta(x) < 0 \) and so, from (8), we get \( \varphi_i(v(x)) < -1/2 \|v(x)\|^2 < 0 \), which in turn implies that \( JF(x)v(x) < 0 \). Whence, by Lemma 2.1, there exists \( \varepsilon > 0 \) such that
\[
F(x + tv(x)) < F(x) + \delta t JF(x)v(x) \quad \text{for all} \ t \in (0, \varepsilon],
\]
and, in particular, \( v(x) \) is a descent direction, i.e., \( F(x + tv(x)) < F(x) \ \forall t \in (0, \varepsilon]. \)

Finally, observe that, while proving Proposition 3.1, we saw that \( v: \mathbb{R}^n \to \mathbb{R}^n \), defined by (7), is bounded on compacts sets of \( \mathbb{R}^n \). Actually, it can be seen something stronger, namely that \( v() \) is a continuous function [28, Lemma 3.3].

3.2 The multiobjective steepest descent algorithm

We are now in conditions to define the extension of the classical steepest descent method for the unconstrained multiobjective optimization problem (1).

**Algorithm 3.2.** The multiobjective steepest descent method (MSDM)

Pesquisa Operacional, Vol. 34(3), 2014
1. Choose \( \sigma \in (0, 1) \) and \( x^0 \in \mathbb{R}^n \). Set \( k := 0 \).

2. Compute \( v^k := v(x^k) = \arg\min_{v \in \mathbb{R}^n} \varphi_{x^k}(v) + \frac{1}{2} \| v \|^2 \),
where \( \varphi_{x^k}(v) := \max_{i=1,\ldots,m} \nabla F_i(x^k)^T v \).

3. Compute \( \theta(x^k) = \varphi_{x^k}(v^k) + \frac{1}{2} \| v^k \|^2 \). If \( \theta(x^k) = 0 \), then stop.

4. Choose \( t_k \) as the largest \( t \in \{1/2^j : j = 0, 1, 2, \ldots \} \) such that
\[
F(x^k + t v^k) \leq F(x^k) + \sigma t JF(x^k)v^k.
\]

5. Set \( x^{k+1} := x^k + t_k v^k \), \( k := k + 1 \) and go to Step 2.

Some comments are in order. Since \( v \mapsto \varphi_{x^k}(v) + \frac{1}{2} \| v \|^2 \) is strongly convex, \( v^k \) exists and is unique. Note that, by Proposition 3.1, the stopping rule \( \theta(x^k) = 0 \) can be replaced by \( v^k = 0 \).

Also note that, if at iteration \( k \) the algorithm does not reach Step 4, i.e., if it stops at Step 3, then, by Proposition 3.1, \( x^k \) is a stationary point for \( F \).

If Step 4 is reached at iteration \( k \), by our comments after Proposition 3.1 and Lemma 2.1, the computation of the step length in Step 4 ends up in a finite number of half reductions. Observe that instead of the factor \( 1/2 \) in the reductions of Step 4 we could have used any other scalar \( \nu \in (0, 1) \).

If the algorithm does not stop, from Step 4 and the fact that \( JF(x^k)v^k < 0 \), we have that the objective values sequence \( \{F(x^k)\} \) is \( \mathbb{R}^m \)-decreasing, i.e.,
\[
F(x^{k+1}) < F(x^k) \quad \text{for all } k.
\]

If the algorithm stops at iteration \( k_0 \), the above inequality holds for \( k = 0, 1, \ldots, k_0 - 1 \).

### 3.3 Convergence analysis of the MSDM: the general case

Since Algorithm 3.2 ends up with a stationary point or produces infinitely many iterates, from now on, we assume that an infinite sequence \( \{x^k\} \) of nonstationary points is generated. We begin presenting a simple application of the standard convergence argument for the scalar-valued steepest descent method.

**Theorem 3.3.** Let \( \{x^k\} \) be a sequence produced by Algorithm 3.2. Every accumulation point of \( \{x^k\} \), if any, is a stationary point.

**Proof.** Let \( \bar{x} \) be an accumulation point of the sequence \( \{x^k\} \). Consider \( v(\bar{x}) \) and \( \theta(\bar{x}) \), given by (7) and (8), respectively, i.e., the solution and the optimal value of (6) at \( x = \bar{x} \). According to Proposition 3.1, it is enough to prove that \( \theta(\bar{x}) = 0 \).

We know that the sequence \( \{F(x^k)\} \) is \( \mathbb{R}^m \)-decreasing (i.e., componentwise strictly decreasing), so we have that
\[
\lim_{k \to \infty} F(x^k) = F(\bar{x}).
\]
Therefore,
\[ \lim_{k \to \infty} \| F(x^k) - F(x^{k+1}) \| = 0. \]
But, by Steps 2-5 of Algorithm 3.2,
\[ F(x^k) - F(x^{k+1}) \geq -t_k \beta JF(x^k)v^k \geq 0, \]
and therefore
\[ \lim_{k \to \infty} t_k JF(x^k)v^k = 0, \]
which, componentwise, can be written as
\[ \lim_{k \to \infty} t_k \nabla F_i(x^k)^T v^k = 0 \text{ for all } i = 1, \ldots, m. \quad (10) \]
Considering that \( t_k \in (0, 1] \) for all \( k \), we have the following two possibilities:
\[ \limsup_{k \to \infty} t_k > 0 \quad (11) \]
or
\[ \limsup_{k \to \infty} t_k = 0. \quad (12) \]
First assume that (11) holds. Then there exists a subsequence \( \{ x^{k_j} \} \) converging to \( \bar{x} \) and \( \bar{t} > 0 \) such that \( \lim_{j \to \infty} t_{k_j} = \bar{t} \). So, using (10), we get
\[ 0 = \lim_{j \to \infty} \nabla F_i(x^{k_j})^T v^{k_j} \leq \lim_{j \to \infty} \psi_{A_k}(v^{k_j}) + \frac{1}{2} \| v^{k_j} \|^2 = \lim_{j \to \infty} \theta(x^{k_j}) = \theta(\bar{x}), \]
where the last equality is a consequence of Proposition 3.1, 2. So, by Proposition 3.1, 3, we conclude that \( \theta(\bar{x}) = 0 \), and \( \bar{x} \) is stationary.
Now assume that (12) holds. Due to the the fact that \( v(\cdot) \) is bounded on compacts, \( v^k = v(x^k) \) for all \( k \) and \( \{ x^k \} \) has a convergent subsequence, it follows that the sequence \( \{ v^k \} \) has also a bounded subsequence. Therefore there exists \( \bar{v} \in \mathbb{R}^m \) and a subsequence \( \{ v^{k_j} \} \) such that \( \lim_{j \to \infty} v^{k_j} = \bar{v} \) and \( \lim_{j \to \infty} t_{k_j} = 0 \). Note that we have
\[ \max_i \nabla F_i(x^{k_j})^T v^{k_j} \leq \theta(v^{k_j}) < 0 \quad \text{for all } j, \]
so letting \( j \to \infty \), we get
\[ \max_i \nabla F_i(\bar{x})^T \bar{v} \leq \theta(\bar{x}) \leq 0. \quad (13) \]
Take now a fixed but arbitrary positive integer \( q \). Since \( t_{k_j} \to 0 \), for \( j \) large enough, we have
\[ t_{k_j} < \frac{1}{2^q}, \]
which means that the Armijo-like condition at \( x^{k_j} \) in Step 4 of Algorithm 3.2 is not satisfied for \( t = 1/2^q \), i.e.,
\[ F(x^j) + (1/2^q)v^{k_j} \nless F(x^{k_j}) + \sigma (1/2^q) JF(x^{k_j})v^{k_j}, \]
If \( h_i \) is convex for all \( i \), the above conditions are also sufficient for the optimality of \( \tilde{x} \) with optimal solution \( \tilde{x} \).

Proof. First note that the problem \( \min_{\tau \in \mathbb{R}^n} \max_{i=1, \ldots, m} h_i(x) \) can be reformulated as

\[
\begin{align*}
\text{minimize} & \quad \tau \\
\text{subject to} & \quad h_i(x) - \tau \leq 0, \quad i = 1, \ldots, m.
\end{align*}
\]

with optimal solution \( (\tilde{\tau}, \tilde{x}) \in \mathbb{R} \times \mathbb{R}^n \). Since the Lagrangian of this problem is given by \( L((\tau, x), \alpha) := \tau + \sum_{i=1}^m \alpha_i (h_i(x) - \tau) \), its Karush-Kuhn-Tucker (KKT) conditions at \( (\tilde{\tau}, \tilde{x}) \) are...
By (6)-(7), \( v(\hat{x}) = \arg\min_{x} \max_{i} \nabla F_i(x)^\top v + (1/2)\|v\|^2 \). Therefore, in view of the above proposition, there exists \( \omega(x) := \alpha(v(x)) \in \mathbb{R}^m_+ \), with \( \sum_{i=1}^m \omega_i(x) = 1 \) such that

\[
\sum_{i=1}^m \omega_i(x) \nabla F_i(x) + v(x) = 0,
\]

and so

\[
v(x) = - \sum_{i=1}^m \omega_i(x) \nabla F_i(x).
\]

The above equality can be rewritten as

\[
v(x) = - J F(x)^\top w(x).
\]

Note that (15) tells us that the steepest descent direction \( v(x) \) is a classical steepest descent direction for the scalar-valued problem \( \min_x \sum_{i=1}^m \omega_i F_i(x) \), whose weights \( \omega_i := \omega_i(x) \geq 0 \) are the \( \alpha \)-priori unknown KKT multipliers for problem (14), with \( h_i = F_i \) for all \( i \), at \( (\theta(x), v(x)) \).

Let us now introduce an important concept. We say that the mapping \( F : \mathbb{R}^m \to \mathbb{R}^m \) is \( \mathbb{R}^m_+ \)-convex if

\[
F(\lambda x + (1 - \lambda) z) \leq \lambda F(x) + (1 - \lambda) F(z) \quad \text{for all } x, z \in \mathbb{R}^m \text{ and all } \lambda \in [0, 1].
\]

Clearly, \( F : \mathbb{R}^m \to \mathbb{R}^m \) is \( \mathbb{R}^m_+ \)-convex if and only if its components \( F_i : \mathbb{R}^m \to \mathbb{R} \) are all convex. Under this additional assumption, we have the following extension of the well-known scalar inequality which establishes that a convex differentiable function always overestimates its linear approximation:

\[
F(u) \geq F(w) + J F(w)(u - w) \quad \text{for any } u, w \in \mathbb{R}^m.
\]

In fact, this vector inequality is equivalent to the following \( m \) scalar inequalities \( F_i(u) \geq F_i(w) + \nabla F_i(w)^\top (u - w) \) for all \( u, w \in \mathbb{R}^m \) and all \( i \).

Let us now state and prove a simple result relating Pareto optimality and stationarity.

**Lemma 3.5.** Let \( \hat{x} \) be a weak Pareto optimum for \( F \). Then \( \hat{x} \) is a stationary point. If, in addition, \( F \) is \( \mathbb{R}^m_+ \)-convex, this condition is also sufficient for weak Pareto optimality. So, under \( \mathbb{R}^m_+ \)-convexity, the concepts of weak efficiency and stationarity are equivalent.

**Proof.** First, assume that \( \hat{x} \) is weak Pareto. As we know, if \( \hat{x} \) is not a stationary point, there exists \( v \in \mathbb{R}^m \) such that \( J F(\hat{x}) v < 0 \), so (3) holds for \( x = \hat{x} \), in contradiction with the weak Pareto optimality of \( \hat{x} \).

Pesquisa Operacional, Vol. 34(3), 2014
Now suppose that \( \hat{x} \) is stationary for the \( \mathbb{R}^n \)-convex mapping \( F \) and take any \( x \in \mathbb{R}^n \). Since \( \hat{x} \) is stationary, (2) holds for \( x = \hat{x}, v := x - \hat{x} \) and some \( j = j(\hat{x}) \). Therefore, by (18) for \( u = x \) and \( w = \hat{x} \), \( F_j(x) \geq F_j(\hat{x}) \), and so \( \hat{x} \) is weak a Pareto optimum for \( F \). □

We will keep on studying the case in which the algorithm does not have finite termination and therefore it produces infinite sequences \( \{x^k\}, \{v^k\} \) and \( \{t_k\} \). Let us now state the additional assumption under which, in the \( \mathbb{R}^m \)-convex case, we will prove full convergence of \( \{x^k\} \) to a stationary point or, in view of the above lemma, to a weak unconstrained optimum of \( F \).

Assumption 3.6. Every \( \mathbb{R}^m \)-decreasing sequence \( \{y^k\} \subseteq F(\mathbb{R}^m) := \{F(x) \in \mathbb{R}^m: x \in \mathbb{R}^n \} \) is \( \mathbb{R}^m \)-bounded from below by a point in \( F(\mathbb{R}^m) \), i.e., for any \( \{y^k\} \) contained in \( F(\mathbb{R}^m) \) with \( y^{k+1} \leq y^k \) for all \( k \), there exists \( \hat{x} \in \mathbb{R}^n \) such that \( F(\hat{x}) \leq y^k \) for all \( k \).

Some comments concerning the generality/restrictiveness of this assumption are in order. In the classical unconstrained (convex) optimization case, this condition is equivalent to existence of solutions of the optimization problem. This assumption, known as \( \mathbb{R}^m \)-completeness, is standard for ensuring existence of efficient points for vector optimization problems [35, Section 3].

We will need the following technical result in order to prove that the MSDM is convergent.

Lemma 3.7. Suppose that \( F \) is \( \mathbb{R}^n \)-convex and let \( \hat{x} \in \mathbb{R}^n \) be such that \( F(\hat{x}) \leq F(x^k) \) for some \( k \). Then, we have

\[
\|\hat{x} - x^{k+1}\|^2 \leq \|\hat{x} - x^k\|^2 + \|x^k - x^{k+1}\|^2.
\]

Proof. By (16), there exists \( w^k := w(x^k) \in \mathbb{R}^m \) such that

\[
v^k = -JF(x^k)^\top w^k.
\]

Using the \( \mathbb{R}^m \)-convexity of \( F \), we have that (18) holds, so \( F(x^k) + JF(x^k)(\hat{x} - x^k) \leq F(\hat{x}) \).

Since \( F(\hat{x}) \leq F(x^k) \), we get

\[
JF(x^k)(\hat{x} - x^k) \leq 0.
\]

Taking into account that \( w^k \in \mathbb{R}^m \) and using the above vector inequality as well as (19), we get

\[
-(v^k)^\top (\hat{x} - x^k) = (w^k)^\top JF(x^k)(\hat{x} - x^k) \leq 0.
\]

Recall that \( x^{k+1} = x^k + t_k w^k \), with \( t_k > 0 \). Therefore, from the above inequality we get

\[
(x^k - x^{k+1})^\top (\hat{x} - x^k) \leq 0,
\]

which implies the desired inequality, because

\[
\|\hat{x} - x^{k+1}\|^2 = \|\hat{x} - x^k\|^2 + \|x^k - x^{k+1}\|^2 + 2(x^k - x^{k+1})^\top (\hat{x} - x^k).\]

\[
\]

We still need a couple of technical results. The first one, basically, establishes that any accumulation point of a sequence produced by the MSDM furnishes an \( \mathbb{R}^m \)-lower bound for the whole objective values sequence. We state it for an arbitrary sequence of nonstationary points, because we will also need it for the convergence analysis of another method.
Lemma 3.8. Let \( \{x^k\} \) be any infinite sequence of nonstationary points in \( \mathbb{R}^n \) such that \( F(x^{k+1}) \leq F(x^k) \) for all \( k \).

1. If \( \hat{x} \) is an accumulation point of \( \{x^k\} \), then
\[
F(\hat{x}) \leq F(x^k) \quad \text{for all} \ k
\]
and
\[
\lim_{k \to \infty} F(x^k) = F(\hat{x}).
\]

Also, \( F \) is constant in the set of accumulation points of \( \{x^k\} \).

2. If \( \{x^k\} \), generated by Algorithm 3.2, has an accumulation point, then all results of item 1 hold for this sequence.

Proof. 1. Let \( \{x^{k_j}\} \) be a subsequence that converges to \( \hat{x} \). Take a fixed but arbitrary \( k \). For \( j \) large enough, we have \( k_j > k \) and \( F(x^{k_j}) \leq F(x^k) \). So letting \( j \to \infty \), we get
\[
F(\hat{x}) \leq F(x^k).
\]

Since the above inequality holds componentwise and \( F_j(x^{k_j}) \to F_j(\hat{x}) \) as \( j \to \infty \) for all \( i \), we have \( \lim_{k \to \infty} F(x^k) = F(\hat{x}) \).

Now let \( \tilde{x} \) be another accumulation point of \( \{x^k\} \). Then there exists a subsequence \( \{x^{k_\ell}\} \) that converges to \( \hat{x} \). By (20), \( F(\tilde{x}) \leq F(x^{k_\ell}) \) for all \( \ell \), so letting \( \ell \to \infty \), we get \( F(\tilde{x}) \leq F(\hat{x}) \). Interchanging the roles of \( \tilde{x} \) and \( \hat{x} \), by the same reasoning, we get \( F(\tilde{x}) \leq F(\hat{x}) \). These two \( \mathbb{R}^m_i \)-inequalities imply that \( F(\tilde{x}) = F(\hat{x}) \).

2. It suffices to note that any (infinite) sequence \( \{x^k\} \) generated by Algorithm 3.2 satisfies \( F(x^{k+1}) < F(x^k) \) for all \( k \).

We now study the speed of convergence to zero of the sequences \( \{x^k\} \) and \( \{\theta(x^k)\} \).

Lemma 3.9. Suppose that \( \{F(x^k)\} \) is \( \mathbb{R}^m_+ \)-bounded from below by \( \tilde{y} \). Then
\[
\sum_{k=0}^{\infty} t_k |\theta(x^k)| < \infty \quad \text{and} \quad \sum_{k=0}^{\infty} t_k \|x^k\|^2 < \infty.
\]

Proof. We know that \( F_i(x^{k+1}) \leq F_i(x^k) + \sigma t_k \nabla F_i(x^k)^\top x^k \) for any \( i \) and all \( k \). Adding up from \( k = 0 \) to \( k = N \), we get
\[
F(x^{N+1}) \leq F(x^0) + \sigma \sum_{k=0}^{N} t_k \nabla F_i(x^k)^\top x^k
\]
\[
\leq F(x^0) + \sigma \sum_{k=0}^{N} t_i \psi_i(x^k)
\]
\[
= F(x^0) + \sigma \sum_{k=0}^{N} t_k \left( \theta(x^k) - \frac{1}{2} \|x^k\|^2 \right).
\]

Pesquisa Operacional, Vol. 34(3), 2014
where we used (4) and (8) for $x = x^k$. As $\bar{y}_i \leq F_i(x^k)$ and $\theta(x^k) < 0$ for all $i$, $k$ (Proposition 3.1, 3), we obtain

$$\sum_{k=0}^{N} \nu_k \left( \theta(x^k) + \frac{1}{2} \|d^k\|^2 \right) \leq \frac{1}{\sigma} \left( F(x^0) - \bar{y}_i \right).$$

Since the last inequality holds for any nonnegative integer $N$, the conclusions follow immediately. □

Before stating and proving the last theorem of this section, let us introduce the main tool for the convergence analysis. Recall that a sequence $\{u^k\} \subset \mathbb{R}^n$ is quasi-Fejér convergent to a set $U \subset \mathbb{R}^n$ if for every $u \in U$ there exists a sequence $\{\varepsilon_k\} \subset \mathbb{R}$, $\varepsilon_k \geq 0$ for all $k$ and such that

$$\|u^{k+1} - u\|^2 \leq \|u^k - u\|^2 + \varepsilon_k$$

for all $k = 0, 1, \ldots$, with $\sum_{k=0}^{\infty} \varepsilon_k < \infty$.

We will need the following well-known result concerning quasi-Fejér convergent sequences, whose proof can be found in [7].

**Theorem 3.10.** If a sequence $\{u^k\}$ is quasi-Fejér convergent to a nonempty set $U \subset \mathbb{R}^n$, then $\{u^k\}$ is bounded. If, furthermore, $\{u^k\}$ has a cluster point $u$ which belongs to $U$, then $\lim_{k \to \infty} u^k = u$.

In [7] it is proved that the steepest descent method for smooth (scalar) convex minimization, with step size obtained using the Armijo rule implemented by a backtracking procedure, is globally convergent to a solution (essentially, under the sole assumption of existence of optima). Using the same techniques, we extend this result to the multiobjective setting, that is to say, we show that any sequence produced by the MSDM converges to an optimum of problem (1), no matter how poor the initial guess might be.

**Theorem 3.11.** Suppose that $F$ is $\mathbb{R}^m_+$-convex and that Assumption 3.6 holds. Then any sequence $\{x^k\}$ produced by Algorithm 3.2 converges to a weak Pareto optimal point $x^* \in \mathbb{R}^n$.

**Proof.** As $\{F(x^k)\}$ is an $\mathbb{R}^m_+$-decreasing sequence, by Assumption 3.6, there exists $\hat{x} \in \mathbb{R}^n$ such that

$$F(\hat{x}) \leq F(x^k) \quad \text{for all } k = 0, 1, \ldots$$

(21)

Now observe that $0 < \nu_k \leq 1$ for all $k$, so

$$\|x^{k+1} - x^k\|^2 \leq \frac{1}{\nu_k} \|x^{k+1} - x^k\|^2 = \nu_k \|x^k\|^2 \quad \text{for all } k = 0, 1, \ldots$$

Therefore, from (21), Lemma 3.9 and the above inequality, it follows that

$$\sum_{k=0}^{\infty} \|x^k - x^{k+1}\|^2 < \infty.$$  (22)

Define

$$L := \{ x \in \mathbb{R}^n : F(x) \leq F(x^k), \ k = 0, 1, \ldots \}.$$
Using the $\mathbb{R}^n_+$-convexity of $F$ and (21), from Lemma 3.7, we see that for any $x \in L$
\[ \|x - x^{k+1}\|^2 \leq \|x - x^k\|^2 + \|x^k - x^{k+1}\|^2 \] for all $k = 0, 1, \ldots$

As $L$ is nonempty, because $\hat{x} \in L$, from (22) and the above inequality, it follows that \{\textit{x^k}\} is quasi-Fejér convergent to the set $L$. Therefore, \{\textit{x^k}\} has accumulation points. Let $x^\ast$ be one of them. By Lemma 3.8, $x^\ast \in L$. Then, using Theorem 3.10, we see that the whole sequence \{\textit{x^k}\} converges to $x^\ast$. We finish the proof by observing that Theorem 3.3 guarantees that $x^\ast$ is stationary and so, from the $\mathbb{R}^n_+$-convexity of $F$, by virtue of Lemma 3.5, $x^\ast$ is a weak Pareto optimum.

\section{The MSDM with relative errors}

Here we briefly sketch an inexact version of the MSDM. Instead of computing $v^k = v(x^k) = \arg \min_{v \in \mathbb{R}^n} \psi_h(v) + \frac{1}{2}\|v\|^2$ in Step 2 of Algorithm 3.2, we accept approximate solutions of the problem with some prespecified tolerance. For a nonstationary point $x \in \mathbb{R}^n$ and a given tolerance $\epsilon \in [0, 1)$, we say that a direction $v \in \mathbb{R}^n$ is an $\epsilon$-approximate solution of $\min_{v \in \mathbb{R}^n} \psi_h(v) + \frac{1}{2}\|v\|^2$ if

\[ \psi_h(v) + \frac{1}{2}\|v\|^2 \leq (1 - \epsilon)\theta(x), \]

where $\psi_h(v) = \max_{i = 1, \ldots, m} \nabla F_i(x)\top v$ and $\theta(x)$ is the optimal value of the above optimization problem. Calling $h_\epsilon$ the objective function in this problem, i.e., $h_\epsilon(v) := \psi_h(v) + \frac{1}{2}\|v\|^2$, we see that $v$ is an $\epsilon$-approximate solution if, and only if, the relative error of $h_\epsilon$ at $v$ is at most $\epsilon$, that is to say,

\[ \frac{|h_\epsilon(v) - h_\epsilon(v(x))|}{h_\epsilon(v(x))} = \left| \frac{h_\epsilon(v) - \theta(x)}{\theta(x)} \right| \leq \epsilon. \]

Clearly, the exact direction $v(x)$ is always $\epsilon$-approximate at $x$ for any $\epsilon \in [0, 1)$. Moreover, $v(x)$ is the unique 0-approximate direction at $x$. Also note that given a nonstationary point $x \in \mathbb{R}^n$ and $\epsilon \in (0, 1)$, an $\epsilon$-approximate direction $v$ is always a descent direction. Indeed, from Proposition 3.1 and the definition of $\epsilon$-approximate solution, it follows that $\nabla F_i(x)\top v \leq \psi_h(v) + \frac{1}{2}\|v\|^2 \leq (1 - \epsilon)\theta(x) < 0$ for all $i$. Whence, $JF_i(x)v < 0$ and so, by Lemma 2.1, $v$ is a descent direction for $F$ at $x$.

We can now define the inexact MSDM in almost the same way as its exact counterpart: in Step 1, there is also the choice of the tolerance parameter $\epsilon \in [0, 1)$, in Step 2, $v^k$ is taken as an $\epsilon$-approximate solution of $\min_{v \in \mathbb{R}^n} \psi_h(v) + \frac{1}{2}\|v\|^2$ and, finally, the stopping criterion in Step 3 becomes $h_\epsilon(v^k) = 0$. In [19, 28] it is shown that the method is well-defined and that it has the same convergence features as the exact one.

\section{The Multiobjective Projected Gradient Method}

The scheme of this section is similar to the previous one. Here, for the constrained multiobjective problem, we propose an extension of the classical scalar-valued projected gradient method. First, we prove some technical results. Then, we present the convergence analysis of the method. And,
finally, we sketch a numerically robust version of the method which admits relative errors on computing the search directions at each iteration.

Let \( F : \mathbb{R}^n \to \mathbb{R}^m \) be a continuously differentiable mapping and \( C \subseteq \mathbb{R}^n \) be a closed and convex set. We are now interested in the following constrained multiobjective optimization problem:

\[
\begin{align*}
\text{minimize} \quad & F(x) \\
\text{subject to} \quad & x \in C.
\end{align*}
\]  

(23)

For this problem, we seek a Pareto optimum (or efficient point), i.e., a feasible point \( (x^* \in C) \) such that there does not exist \( x \in C \) with \( F(x) \leq F(x^*) \) and \( F(x) \neq F(x^*) \), or we look for a weak Pareto optimum (or weakly efficient point), that is to say, a feasible \( x^* \) such that there does not exist \( x \in C \) with \( F(x) < F(x^*) \).

4.1 The search direction for the multiobjective projected gradient method

For the constrained multiobjective problem, a necessary condition for optimality of a point \( \bar{x} \in C \) is stationarity (or criticality):

\[
J_F(\bar{x})(C - \bar{x}) \cap [-\mathbb{R}^m_+] = \emptyset,
\]

where \( J_F(\bar{x})(C - \bar{x}) := \{ J_F(\bar{x})(x - \bar{x}) : x \in C \} \). So \( \bar{x} \in C \) is stationary for \( F \) if, and only if, for all \( v \in C - \bar{x} \) we have \( J_F(\bar{x})v \neq 0 \). Note that when \( m = 1 \), we retrieve the classical stationarity condition for constrained scalar-valued optimization: \( \langle \nabla F(\bar{x}), x - \bar{x} \rangle \geq 0 \) for all \( x \in C \).

Let us now develop an extension of the classical (scalar) projected gradient method for the vector-valued problem (23). First, for \( x \in C \), we extend the search direction, originally defined for \( C = \mathbb{R}^n \) in (7), by

\[
v(x) := \arg\min_{v \in C - x} \beta \phi_x(v) + \frac{1}{2} \|v\|^2,
\]  

(24)

where \( \beta > 0 \) is a parameter and \( \phi_x \) is defined in (4). Note that, in view of the strong convexity of the minimand, for any \( x \in C \), the projected gradient direction \( v(x) \) is always well-defined.

Now we extend the optimal value function \( \theta(\cdot) \) (given by (8) in the unconstrained case):

\[
\theta(x) := \min_{v \in C - x} \beta \phi_x(v) + \frac{1}{2} \|v\|^2 = \beta \phi_x(v(x)) + \frac{1}{2} \|v(x)\|^2.
\]  

(25)

Proposition 4.1. Let \( v : C \to \mathbb{R}^n \) and \( \theta : C \to \mathbb{R} \) be given by (24) and (25), respectively. Then, the following statements hold.

1. \( \theta(x) \leq 0 \) for all \( x \in C \).
2. The function \( \theta(\cdot) \) is continuous.
3. The following conditions are equivalent:
(a) The point \( x \in C \) is nonstationary.

(b) \( \theta(x) < 0 \).

(c) \( v(x) \neq 0 \).

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

Proof. 1. Recalling that \( 0 \in C - x \), the result follows as in the proof of Proposition 3.1.

2. Let \( x \in C \) and \( \{x^{(i)}\} \subset C \) be a sequence such that \( \lim_{k \to \infty} x^{(i)} = x \). Since \( v(x) \in C - x \), we have \( v(x) + x - x^{(i)} \in C - x^{(i)} \). Thus, from (24) and (25), we have

\[
\theta(x^{(i)}) \leq \beta \varphi_{\rho,0}(v(x) + x - x^{(i)}) + \frac{1}{2} v(x) + x - x^{(i)}\|^2.
\]

Hence, from the first inequality of (5), we obtain

\[
\theta(x^{(i)}) \leq \beta \varphi_{\rho,0}(v(x)) + \beta \varphi_{\rho,0}(x - x^{(i)}) + \frac{1}{2} \|v(x)\|^2 + \frac{1}{2} \|x - x^{(k)}\|^2 + \langle v(x), x - x^{(k)} \rangle.
\]

Since \( F \) is continuously differentiable and \( u \mapsto \max [u_i] \) is continuous, taking \( \limsup_{k \to \infty} \) on both sides of the above inequality yields

\[
\limsup_{k \to \infty} \theta(x^{(i)}) \leq \theta(x).
\]

(26)

Now, observe that since \( v(x^{(k)}) \in C - x^{(k)} \), we have \( v(x^{(k)}) + x^{(k)} - x \in C - x \). Once again from (24) and (25), we obtain

\[
\theta(x) \leq \beta \varphi_{\rho,0}(v(x^{(k)}) + x^{(k)} - x) + \frac{1}{2} \|v(x^{(k)}) + x^{(k)} - x\|^2.
\]

Using once again the first inequality of (5), we have

\[
\theta(x) \leq \beta \varphi_{\rho,0}(v(x^{(k)})) + \beta \varphi_{\rho,0}(x^{(k)} - x) + \frac{1}{2} \|v(x^{(k)})\|^2 + \frac{1}{2} \|x^{(k)} - x\|^2 + \langle v(x^{(k)}), x^{(k)} - x \rangle.
\]

Note that for all \( i \) and any \( z \in C \), from the Cauchy-Schwarz inequality and item 1, we have that

\[ -\|\nabla F(z)\| \|v(z)\| + \frac{1}{2} \|z\|^2 \leq \theta(z) \leq 0. \]

Therefore, since \( F \) is continuously differentiable and \( \{x^{(k)}\} \) is bounded, because it converges, we see that \( \|v(x^{(k)})\| \) is bounded for all \( k \). So, taking \( \liminf_{k \to \infty} \) on both sides of the above inequality, we get

\[
\theta(x) \leq \liminf_{k \to \infty} \beta \varphi_{\rho,0}(v(x^{(k)})) + \|v(x^{(k)})\|/2
\]
\[ = \liminf_{k \to \infty} \beta \varphi_{\rho,0}(v(x^{(k)})) \]
\[ = \liminf_{k \to \infty} \beta \|F(x) - JF(x^{(k)})\| \|v(x^{(k)})\| \]
\[ \leq \liminf_{k \to \infty} \theta(x^{(k)}), \]

Pesquisa Operacional, Vol. 34(3), 2014
where the second inequality follows from (5). We complete the proof by combining this inequality with (26).

3. All results follow as in the proof of Proposition 3.1. (A formal proof of this equivalence for a more general case can be found in [25, Proposition 3].) □

4.2 The multiobjective projected gradient algorithm

We can now define the extension of the classical (scalar) projected gradient method for the constrained multiobjective optimization problem (23).

Algorithm 4.2. The multiobjective projected gradient method (MPGM)

1. Choose $\beta > 0$, $\sigma \in (0, 1)$ and $x^0 \in C$. Set $k := 0$.
2. Compute $v^k := v(x^k) = \arg\min_{v \in C - x^k} \beta \phi_{x^k}(v) + \frac{1}{2}\|v\|^2$, where\[\phi_{x^k}(v) := \max_{i=1,...,m} \nabla F_i(x^k)^\top v.\]
3. Compute $\theta(x^k) = \beta \phi_{x^k}(v^k) + \frac{1}{2}\|v^k\|^2$. If $\theta(x^k) = 0$, then stop.
4. Choose $t_k$ as the largest $t \in \{1/2^j : j = 0, 1, 2, \ldots\}$ such that\[F(x^k + tv^k) \leq F(x^k) + \sigma t JF(x^k)v^k.\]
5. Set $x^{k+1} := x^k + t_kv^k$, $k := k + 1$ and go to Step 2.

Note that, whenever $C = \mathbb{R}^n$, taking $\beta = 1$, we retrieve the MADM. Observe that, by virtue of Proposition 4.1, an alternative stopping criterion in Step 3 is $v^k = 0$. Note also that Algorithm 4.2 either terminates in a finite number of iterations with a stationary point or it generates an infinite sequence of nonstationary points (Proposition 4.1).

Now observe that if Step 4 is reached, by virtue of Lemma 2.1, in finitely many half reductions, we obtain the step length which is used in Step 5 to define the next iterate. Note that instead of the factor $1/2$ in the reductions of Step 4, we could have used any other scalar $\nu \in (0, 1)$. Finally, observe that since $\theta(x^k) < 0$ implies $JF(x^k)v^k < 0$, from Steps 3-5, it follows that whenever the method generates and infinite sequence, the objective values are $\mathbb{R}_m^+$-decreasing, i.e.,

\[F(x^{k+1}) < F(x^k)\quad \text{for all } k = 0, 1, \ldots\]

If the method stops at iteration $k_0$, this inequality hold for $k = 0, 1, \ldots, k_0 - 1$.

4.3 Convergence analysis of the MPGM: the general case

As we know, Algorithm 4.2 has finite termination, ending with a stationary point, or generates an infinite sequence of nonstationary iterates. In the sequel we study the second case. First, we show a simple fact: the algorithm produces feasible sequences.
Lemma 4.3. Let \( \{x^k\} \subset \mathbb{R}^n \) be a sequence generated by Algorithm 4.2. Then, we have \( \{x^k\} \subset C \).

Proof. Let us use an inductive argument. From the initialization of the algorithm, \( x^0 \in C \). Now assume that \( x^k \in C \). Observe that, by Step 4 of Algorithm 4.2, \( t_k \in (0, 1] \) and, by Step 2, \( v^k \in C - x^k \). Then, for some \( z^k \in C \), \( v^k = z^k - x^k \) and from the convexity of \( C \), it follows that \( x^{k+1} = x^k + t_k v^k = (1 - t_k)x^k + t_k z^k \in C \) and the proof is complete. \( \square \)

Now we see that whenever a sequence produced by Algorithm 4.2 has accumulation points, they are all feasible stationary points for the problem.

Theorem 4.4. Every accumulation point, if any, of a sequence \( \{x^k\} \) generated by Algorithm 4.2 is a feasible stationary point.

Proof. The feasibility follows combining the fact that \( C \) is closed with Lemma 4.3. The rest of the proof is similar to the one of Theorem 3.3: it suffices to use the definitions of \( v(x) \) and \( \theta(x) \) given in (24) and (25), respectively. \( \square \)

4.4 Full convergence of the MPMG: the \( \mathbb{R}^m_+ \)-convex case

Let us note that, as being the maximum of convex functions, the objective of the minimization problem which defines \( v(x) \) in (24) is also convex and, as we also have that \( C - x \) is a closed convex subset of \( \mathbb{R}^n \), by [3, Proposition 4.7.2], a necessary condition for the optimality of \( v(x) \) is the existence of \( w(x) \in \mathbb{R}^m \) with

\[
\begin{align*}
w_i(x) &\geq 0 \quad \text{and} \quad \sum_{i=1}^{m} w_i(x) = 1 \\
\beta \sum_{i=1}^{m} w_i(x) \nabla F_i(x) + v(x), v - v(x) \end{align*}
\]

such that

\[
\left\{ \beta \sum_{i=1}^{m} w_i(x) \nabla F_i(x) + v(x), v - v(x) \right\} \geq 0 \quad \text{for all } v \in C - x. \tag{28}
\]

This condition can be written as

\[
\left( x - \beta \sum_{i=1}^{m} w_i(x) \nabla F_i(x) - (x + v(x)), u - (x + v(x)) \right) \leq 0 \quad \text{for all } u \in C.
\]

Now using a very well-known characterization of orthogonal projections (see, for instance, [3, Proposition 2.2.1]), we have

\[
x + v(x) = P_C \left( x - \beta \sum_{i=1}^{m} w_i(x) \nabla F_i(x) \right),
\]

where \( P_C(z) := \inf \{ \|u - z\| : u \in C \} \) is well-defined because \( C \) is closed and convex. We conclude that there exists \( w(x) \in \mathbb{R}^m \) satisfying (27) such that

\[
v(x) = P_C \left( x - \beta J F(x)^\top w(x) \right) - x. \tag{29}
\]
Observe that in the constrained scalar-valued case (m = 1), we have \( JF(x)^\top = \nabla F(x) \) and \( w(x) = 1 \), which means that

\[
v(x) = P_C(x - \beta \nabla F(x)) - x,
\]

and this is precisely the search direction used in the projected gradient method for real-valued optimization. So this is an extension to the multiobjective setting of the classical projected gradient method.

**Lemma 4.5.** Suppose that \( F \) is \( \mathbb{R}_+^m \)-convex. Let \( \{x^k\} \) be an infinite sequence generated by Algorithm 4.2. If for \( \tilde{x} \in C \) and \( k \geq 0 \), we have \( F(\tilde{x}) \leq F(x^k) \), then

\[
\|x^{k+1} - \tilde{x}\|^2 \leq \|x^k - \tilde{x}\|^2 + 2\beta t_k \|w^k, JF(x^k)^\top \|^2,
\]

where \( w^k := w(x^k) \in \mathbb{R}^m \) is such that (27) and (29) hold with \( x = x^k \).

**Proof.** Since \( x^{k+1} = x^k + t_k \hat{v}^k \), we have

\[
\|x^{k+1} - \tilde{x}\|^2 = \|x^k - \tilde{x}\|^2 + 2\beta t_k \|w^k, JF(x^k)^\top \|^2 - 2t_k \langle \hat{v}^k, \tilde{x} - x^k \rangle. \tag{30}
\]

Let us analyze the rightmost term of the above expression. Since \( \hat{v}^k = v(x^k) \) and \( w^k = w(x^k) \), by (28) with \( x = x^k \), we get

\[
\langle \beta JF(x^k)^\top w^k + \hat{v}^k, v - \hat{v}^k \rangle \geq 0 \quad \text{for all } v \in C - x^k.
\]

Taking \( v = \tilde{x} - x^k \in C - x^k \) in the above inequality, we obtain

\[
\langle \hat{v}^k, \tilde{x} - x^k \rangle \leq \beta \langle w^k, JF(x^k)(\tilde{x} - x^k) \rangle - \beta \langle w^k, JF(x^k)^\top \rangle - \|\hat{v}^k\|^2. \tag{31}
\]

Now, observe that, the convexity of \( F \) and (18) yield \( JF(x^k)(\tilde{x} - x^k) \leq F(\tilde{x}) - F(x^k) \). This fact, together with \( w^k \geq 0 \), and \( F(\tilde{x}) \leq F(x^k) \) implies

\[
\langle w^k, JF(x^k)(\tilde{x} - x^k) \rangle \leq \langle w^k, F(\tilde{x}) - F(x^k) \rangle \leq 0.
\]

Also, since \( JF(x^k)^\top \hat{v}^k \leq 0 \), because \( x^k \) is nonstationary, we have \( \langle w^k, JF(x^k)^\top \hat{v}^k \rangle < 0 \). Thus, recalling that \( \beta > 0 \), from (31) it follows that

\[
\langle \hat{v}^k, \tilde{x} - x^k \rangle \leq \beta \langle w^k, JF(x^k)^\top \rangle - \|\hat{v}^k\|^2.
\]

The above inequality, together with (30), shows that

\[
\|x^{k+1} - \tilde{x}\|^2 \leq \|x^k - \tilde{x}\|^2 + 2\beta t_k \|w^k, JF(x^k)^\top \|^2 + 2\beta t_k \|w^k, JF(x^k)^\top \| - 2t_k \|\hat{v}^k\|^2.
\]

And the result follows because \( t_k \geq 0 \).

As in the unconstrained case, we need to make an extra assumption in order to prove full convergence of the method.
Assumption 4.6. Every $\mathbb{R}^m_+$-decreasing sequence $\{y^k\} \subset F(C) := \{F(x) : x \in C\}$ is $\mathbb{R}^m_+$-bounded from below by an element of $F(C)$.

We can now state and prove the main result on the convergence analysis of the MPGM: under similar conditions of those of Theorem 3.11, regardless of the initial (feasible) guess, the method converges to an optimum of problem (23).

Theorem 4.7. Assume that $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is $\mathbb{R}^m_+$-convex and that Assumption 4.6 holds. Then, any sequence $\{x^k\}$ generated by Algorithm 4.2 converges to a weak Pareto optimum.

Proof. Let us define the set

$$L := \{x \in C : F(x) \leq F(x^k) \text{ for all } k\},$$

and take $\hat{x} \in L$, which exists by Assumption 4.6. Since $F$ is $\mathbb{R}^m_+$-convex, it follows from Lemma 4.5 that

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\beta_k \|F'(x^k)(v^k)\|^2$$

for all $k$. (32)

First, we prove that the sequence $\{x^k\}$ is quasi-Féjer convergent to the set $L$. Let $e^1, \ldots, e^m$ be the canonical basis of $\mathbb{R}^m$. We observe that, for each $k$, we can write

$$w^k = \sum_{i=1}^m w^k_i e^i.$$

Let us note that, from (27), $0 \leq w^k_i \leq 1$ for all $i$ and $k$. Thus, from (32), we have

$$\|x^{k+1} - \hat{x}\|^2 \leq \|x^k - \hat{x}\|^2 + 2\beta_k \sum_{i=1}^m |\langle e^i, JF(x^k)v^k \rangle|$$

for all $k$.

Defining $\varepsilon_k := 2\beta_k \sum_{i=1}^m |\langle e^i, JF(x^k)v^k \rangle|$, we have that $\varepsilon_k \geq 0$ and

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

Thus, let us prove that $\sum_{k=0}^\infty \varepsilon_k < \infty$. From the Armijo-like condition in Step 4 of the algorithm, we obtain

$$-\varepsilon_k \langle e^i, JF(x^k)v^k \rangle \leq \frac{1}{\sigma} (F_i(x^k) - F_i(x^{k+1})).$$

Since $x^k$ is nonstationary, we also have for all $i$ and $k$ that

$$\langle e^i, JF(x^k)v^k \rangle = \nabla F_i(x^k)^\top v^k < 0,$$

which means that $-\varepsilon_k \langle e^i, JF(x^k)v^k \rangle = \varepsilon_k |\langle e^i, JF(x^k)v^k \rangle|$. Hence, we obtain

$$\varepsilon_k \leq \frac{2\beta_k}{\sigma} \sum_{i=1}^m (F_i(x^k) - F_i(x^{k+1})).$$

Pesquisa Operacional, Vol. 34(3), 2014
Adding up from \( k = 0 \) to \( k = N \) at the above inequality, where \( N \) is any positive integer, we get

\[
\sum_{k=0}^{N} \varepsilon_k \leq \frac{2\beta}{\sigma} \sum_{i=1}^{m} (F_i(x^{0}) - F_i(x^{N+1})).
\]

Since \( N \) is an arbitrary positive integer, \( \beta, \sigma > 0 \) and \( F(\hat{x}) \leq F(x^k) \) for all \( k \), we conclude that \( \sum_{k=0}^{\infty} \varepsilon_k < \infty \). Therefore, recalling that \( \hat{x} \) was an arbitrary element of \( L \), we see that \( x^k \) converges quasi-Féjer to \( L \).

By Theorem 3.10, it follows that \( \{x^k\} \) is bounded. So, since \( C \) is closed, \( \{x^k\} \subseteq C \) has at least one feasible accumulation point, say \( x^* \), which is feasible and stationary by Theorem 4.4. Using now the \( \mathbb{R}^n_+ \)-convexity of \( F \), from Lemma 3.5, it follows that \( x^* \in C \) is a weak Pareto optimum.

As in the end of the proof of Theorem 3.11, we now see that \( x^* \in L \). Since \( F(x^{k+1}) < F(x^k) \) for all \( k \), by Lemma 3.8, 1, \( F(x^k) \leq F(x^*) \) for all \( k \), and \( x^* \in L \) So, once again from Theorem 3.10, we conclude that \( \{x^k\} \) converges to \( x^* \in C \), a weak Pareto optimum. \( \square \)

4.5 The MPGM with relative errors

As we did for the MSDM in Subsection 3.5, we present here a quick overview of an MPGM version implemented with relative errors in the search directions. Given a nonstationary \( x \in C \) and a tolerance \( \varepsilon \in [0, 1) \), we say that \( v \in C - x \) is an \( \varepsilon \)-approximate solution of the problem

\[
\min_{v \in C - x} \psi(x) + \frac{1}{2} \|v\|^2
\]

if

\[
h_\varepsilon(v) \leq (1 - \varepsilon)\theta(x),
\]

where

\[
h_\varepsilon(v) := \psi(x) + 1/2\|v\|^2,
\]

with \( \psi(x) \) given by (4).

Similar comments as those made in Subsection 3.5 for \( \varepsilon \)-approximate solutions for unconstrained problems can be made here for the constrained case. The modification of the (exact) MPGM is basically the same as those proposed in Subsection 3.5 for the (exact) MSDM: in Step 2 of Algorithm 4.2, instead of defining \( v^k \) as the exact solution of \( \min_{v \in C - x} \psi(x) + 1/2\|v\|^2 \), we take \( v^k \) as an \( \varepsilon \)-approximate solution of this problem. The stopping rule at Step 3 \( (\theta(v^k) = 0) \) is replaced by \( h_\varepsilon(v^k) = 0 \).

For a more general case (vector optimization), a comprehensive study of the inexact version of the MPGM can be found in [22], where, under essentially the same hypotheses made here for the (exact) MPGM, all convergence results are fully proved.

5 THE MULTIOBJECTIVE NEWTON METHOD

We now go back to the unconstrained multiobjective problem (1). In order to define a Newton-type search direction for this kind of problems, we first recall some convexity notions for vector-valued mappings. We say that \( F : \mathbb{R}^n \to \mathbb{R}^m \) is strictly \( \mathbb{R}^m_+ \)-convex if the inequality (17) is satisfied strictly, that is, if \( F \) is componentwise strictly convex. We also say that \( F \) is strongly \( \mathbb{R}^m_+ \)-convex if there exists \( \hat{u} \in \mathbb{R}^m_+ \), such that

\[
F(\lambda x + (1 - \alpha)y) \leq \lambda F(x) + (1 - \lambda)F(y) - \frac{1}{2}(1 - \lambda)||x - y||^2 \hat{u}.
\]
for all \( x, y \in \mathbb{R}^m \) and all \( \lambda \in (0, 1) \), which is equivalent to say that \( F \) is componentwise strongly convex (each component \( F_i \) has \( \lambda \), as its modulus of strong convexity). Clearly, as in the scalar case, strong \( \mathbb{R}_+^m \)-convexity implies strict \( \mathbb{R}_+^m \)-convexity, which, in turn, implies \( \mathbb{R}_+^m \)-convexity.

The following result relates Pareto optimality with strict \( \mathbb{R}_+^m \)-convexity and characterizes strong \( \mathbb{R}_+^m \)-convexity in terms of the eigenvalues of the objective functions’ Hessians.

**Lemma 5.1.** Assume that \( F: \mathbb{R}^n \rightarrow \mathbb{R}^m \) is twice continuously differentiable. Then, the following properties hold.

1. If \( F \) is strictly \( \mathbb{R}_+^m \)-convex and \( \bar{x} \in \mathbb{R}^n \) is stationary for \( F \), then \( \bar{x} \) is a Pareto optimal point.
2. The mapping \( F \) is strongly \( \mathbb{R}_+^m \)-convex if and only if there exists \( \rho > 0 \) such that
   \[
   \lambda_{\min}(\nabla^2 F_i(x)) \geq \rho \quad \text{for all } x \in \mathbb{R}^n \text{ and all } i = 1, \ldots, m,
   \]
   where \( \lambda_{\min}: \mathbb{R}^{n \times n} \rightarrow \mathbb{R} \) denotes the smallest eigenvalue function.

**Proof.** It follows from [27, Corollary 2.2 and Proposition 2.3]. \( \square \)

### 5.1 The search direction for the multiobjective Newton method

From now on, we will assume that \( F \) is strongly \( \mathbb{R}_+^m \)-convex and twice continuously differentiable. For all \( i = 1, \ldots, m \), let us define \( \psi_i: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} \) by

\[
\psi_i(x,s) := \nabla F_i(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_i(x)s.
\]

Hence, for any \( i, s \mapsto F_i(x) + \psi_i(x, s) \) is the local quadratic approximation of \( F_i \) at \( x \). For \( x \in \mathbb{R}^n \), we define the Newton direction for \( F \) at \( x \) as

\[
s(x) := \arg\min_{s \in \mathbb{R}^n} \max_{i = 1, \ldots, m} \psi_i(x, s).
\]

Since \( \psi_i(x, \cdot) \) is strongly convex for all \( i \), \( s \mapsto \max_{i = 1, \ldots, m} \psi_i(x, \cdot) \) is also strongly convex and it has a unique minimizer in \( \mathbb{R}^n \), and so \( s(x) \) is well-defined. Clearly, for \( m = 1 \) we retrieve the Newton direction for unconstrained real-valued optimization.

Observe that whenever \( \nabla^2 F_i(x) \) is equal to \( I \), the identity matrix of \( \mathbb{R}^{n \times n} \), for all \( i \), we have that

\[
\max_{i = 1, \ldots, m} \psi_i(x, s) = \psi(x) + (1/2)\|s\|^2,
\]

where \( \psi \) is given by (4). Therefore, in multiobjective optimization, if \( \nabla^2 F_i(x) = I \) for all \( i \), then, as it happens in the scalar case, the Newton and the steepest descent directions coincide.

Let us call \( \theta(x) \) the optimal value of the minimization problem \( \min_{x \in \mathbb{R}^n} \max_{i = 1, \ldots, m} \psi_i(x, s) \), that is to say,

\[
\theta(x) := \max_{i = 1, \ldots, m} \psi_i(x, s(x)).
\]

In order to compute \( s(x) \), we can reformulate the (possibly nonsmooth) min-max problem as:

\[
\text{minimize} \quad \tau \quad \text{subject to} \quad \nabla F_i(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_i(x)s - \tau \leq 0, \quad i = 1, \ldots, m,
\]

Pesquisa Operacional, Vol. 34(3), 2014
a smooth convex problem with variables \((\tau, v) \in \mathbb{R} \times \mathbb{R}^n\) for which there are many efficient numerical solvers.

We now present analogous results as those of Propositions 3.1 and 4.1.

**Proposition 5.2.** Let \(s : \mathbb{R}^n \to \mathbb{R}^n\) and \(\theta : \mathbb{R}^n \to \mathbb{R}\) be given by (34) and (35), respectively. Then, the following statements hold.

1. \(\theta(x) \leq 0\) for all \(x \in \mathbb{R}^n\).
2. The following conditions are equivalent:
   (a) The point \(x\) is nonstationary.
   (b) \(\theta(x) < 0\).
   (c) \(s(x) \neq 0\).

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

**Proof.**

1. By (33) and (35), \(\theta(x) \leq \max_i \psi_i(x, 0) = 0\).
2. (a) \(\Rightarrow\) (b). By (a), there exists \(\tilde{s} \in \mathbb{R}^n\) such that \(JF(x)\tilde{s} \in -\mathbb{R}^{n \times +}\), which means that \(\nabla F_i(x)^\top \tilde{s} < 0\) for all \(i\). From (33) and (35), for \(\tau > 0\), we get
   \[
   \theta(x) \leq \tau \left( \max_i \nabla F_i(x)^\top \tilde{s} + \frac{\tau}{2} \nabla^2 F_i(x) \tilde{s} \right).
   \]

Therefore, for \(\tau > 0\) small enough the right hand side of the above inequality is negative and (b) holds.

(b) \(\Rightarrow\) (c). Note that if \(\theta(x) < 0\) then, in view of (33) and (35), we have \(s(x) \neq 0\).

(c) \(\Rightarrow\) (a). From the strong \(\mathbb{R}^n_+\)-convexity of \(F\), it follows that \(\nabla^2 F_i(x)\) is positive definite for all \(i\), so from (c) and item 1, we get
   \[
   \nabla F_i(x)^\top s(x) < \nabla F_i(x)^\top s(x) + \frac{1}{2} s(x)^\top \nabla^2 F_i(x) s(x) \leq \theta(x) \leq 0 \quad \text{for all } i = 1, \ldots, m.
   \]

Hence, \(JF(x)s(x) < 0\) and (a) holds. \(\square\)

In particular, we just saw that, for any nonstationary \(x \in \mathbb{R}^n\), \(s(x)\) is a descent direction. The next lemma supplies an implicit definition of \(s(x)\), which will be used in order to obtain bounds for \(\|s(x)\|\). These bounds will allow us to prove continuity of \(s(\cdot), \theta(\cdot)\) and will be used to establish the rate of convergence of the Newton method.

**Lemma 5.3.** For any \(x \in \mathbb{R}^n\), the Newton direction defined in (34) can be written as
   \[
   s(x) = -\left[ \sum_{i=1}^m a_i \nabla^2 F_i(x) \right]^{-1} \sum_{i=1}^m a_i \nabla F_i(x),
   \]
   with \(a_i = a_i(x) \geq 0\) and \(\sum_{i=1}^m a_i = 1\).
Proof. We know that the convex problem \( \min_j \max_i \psi_i(x, s) \), where \( \psi_i \) is defined in (33), has a unique solution \( s(x) \). Applying Proposition 3.4, we see that there exists \( a := a(x) \in \mathbb{R}^m \) such that the following conditions hold at \( s = s(x) \) and \( \tau = \theta(x) \):

\[
\sum_{i=1}^{m} a_i = 1, \quad \sum_{i=1}^{m} a_i \left( \nabla F_i(x) + \nabla^2 F_i(x) s \right) = 0, \tag{37}
\]

\[
a_i \geq 0, \quad \nabla F_i(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_i(x) s - \tau \leq 0, \quad i = 1, \ldots, m, \tag{38}
\]

\[
a_i \left( \nabla F_i(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_i(x) s - \tau \right) = 0, \quad i = 1, \ldots, m. \tag{39}
\]

Since \( \nabla^2 F_i(x) \) is positive definite and the multipliers \( a_i \) are nonnegative for all \( i \), from both conditions in (37), the result follows.

So \( s(x) \), defined in (34), is a Newton direction for a standard scalar optimization problem \( (\min_i \sum_{i=1}^{m} a_i F_i(x)) \), implicitly induced by weighting the given objective functions by the non-negative \( \text{a priori} \) unknown KKT multipliers. We now give upper bounds for the norm of the Newton direction.

Corollary 5.4. For any \( x \in \mathbb{R}^n \), we have

\[
\|s(x)\| \leq \frac{2}{\rho} \| J F(x) \| \quad \text{and} \quad \|s(x)\|^2 \leq \frac{2}{\rho} |\theta(x)|,
\]

where \( \rho > 0 \) is given by Lemma 5.1.

Proof. Let us proof the first inequality. By Proposition 5.2, we have \( \theta(x) \leq 0 \), which, by (35), means that \( \max_i \psi_i(x, s(x)) \leq 0 \). Hence, by (33),

\[
\frac{1}{2} s(x)^\top \nabla^2 F_i(x) s(x) \leq -\nabla F_i(x)^\top s(x) = -\{J F(x) s(x)\}_i \quad \text{for all } i.
\]

From Lemma 5.1, 2, we get

\[
\frac{\rho}{2} \|s(x)\|^2 \leq \frac{1}{2} s(x)^\top \nabla^2 F_i(x) s(x) \quad \text{for all } i.
\]

The proof follows from these two inequalities and the operator norm properties.

We now proof the second inequality. We know that the complementarity condition (39) holds for \( (\tau, s) = (\theta(x), s(x)) \). So, by (33) and (35), \( a_i > 0 \) implies \( \theta(x) = \psi_i(x, s(x)) \), and then, using the first equality in (37), we have

\[
\theta(x) = \sum_{i=1}^{m} a_i \psi_i(x, s(x))
\]

\[
= \sum_{i=1}^{m} a_i \nabla F_i(x)^\top s(x) + \frac{1}{2} s(x)^\top \sum_{i=1}^{m} a_i \nabla^2 F_i(x) s(x).
\]
Replacing $\sum_{i=1}^{m} \alpha_i \nabla F_i(x)$ by its expression obtained from (36), using Lemma 5.1, 2 and the fact that $\sum \alpha_i = 1$ with $\alpha_i \geq 0$ for all $i$, we conclude that

$$\theta(x) = -\frac{1}{2}s(x)^\top \sum_{i=1}^{m} \alpha_i \nabla^2 F_i(x)s(x) \leq -\frac{\rho}{2} \|s(x)\|^2,$$

and the proof follows immediately from Proposition 5.2, 1.

We now establish the continuity of the functions $\theta(\cdot)$ and $s(\cdot)$.

**Proposition 5.5.** The functions $\theta: \mathbb{R}^n \to \mathbb{R}$ and $s: \mathbb{R} \to \mathbb{R}^n$, given by (35) and (34) respectively, are continuous.

**Proof.** Take $x, x' \in \mathbb{R}^n$. Let $s = s(x), s' = s(x')$ and $i_0$ such that $\max_i \psi_i(x', s) = \psi_i(x', s)$. Then, from (35), (33) and the Cauchy-Schwarz inequality,

$$\theta(x') \leq \nabla F_{i_0}(x')^\top s + \frac{1}{2} s^\top \nabla^2 F_{i_0}(x') s$$

$$= \nabla F_{i_0}(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_{i_0}(x) s$$

$$+ (\nabla F_{i_0}(x') - \nabla F_{i_0}(x))^\top s + \frac{1}{2} s^\top \nabla^2 F_{i_0}(x') - \nabla^2 F_{i_0}(x)) s$$

$$\leq \theta(x) + \|s\| \|\nabla F(x') - \nabla F(x)\| + \frac{1}{2} \|s\|^2 \|\nabla^2 F(x) - \nabla^2 F(x')\|,$$

where $\nabla^2 F$ stands for the second derivative of $F$. Therefore, from the above inequality and Corollary 5.4 we conclude that

$$\theta(x') - \theta(x) \leq \frac{2}{\rho} \|\nabla F(x') - \nabla F(x)\| \|\nabla^2 F(x) - \nabla^2 F(x')\| + \frac{2}{\rho} \|\nabla^2 F(x)\|^2 \|\nabla^2 F(x) - \nabla^2 F(x')\|.$$

Since the above inequality also holds interchanging $x$ and $x'$ and $F$ is twice continuously differentiable, we conclude that $\theta(x)$ is continuous.

Now we prove the continuity of $s(\cdot)$. By the same reasoning as above,

$$\theta(x) = \max_i \left[ \nabla F_i(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_i(x)s \right]$$

$$= \max_i \left[ \nabla F_i(x')^\top s + \frac{1}{2} s^\top \nabla^2 F_i(x')s \right.$$

$$+ (\nabla F_i(x) - \nabla F_i(x'))^\top s + \frac{1}{2} s^\top (\nabla^2 F_i(x) - \nabla^2 F_i(x')) s \bigg]$$

$$\geq \max_i \left[ \nabla F_i(x')^\top s + \frac{1}{2} s^\top \nabla^2 F_i(x')s \right.$$

$$- \|s\| \|\nabla F(x') - \nabla F(x)\| - \frac{1}{2} \|s\|^2 \|\nabla^2 F(x) - \nabla^2 F(x')\|$$

$$= \max_i \psi_i(x', s) - \|s\| \|\nabla^2 F(x') - \nabla^2 F(x)\| - \frac{1}{2} \|s\|^2 \|\nabla^2 F(x) - \nabla^2 F(x')\|,$$

where we used the Cauchy-Schwarz inequality and (33) with $x = x'$. By Lemma 5.1, 2 $\lambda_{\min}(\nabla^2 F_i(x)) \geq \rho$ for all $x \in \mathbb{R}^n$ and all $i$. Therefore, $\psi_i(x', \cdot)$ is strongly convex with
modulus $\rho$ for all $i$ and so is $\max_i \psi_i(x', \cdot)$. Hence, using the fact that $\max_i \psi_i(x', \cdot)$ is minimized at $s'$, we have

$$\max_i \psi_i(x', s) \geq \max_i \psi_i(x', s') + \frac{\rho}{2} \|s - s'\|^2 = \theta(x') + \frac{\rho}{2} \|s - s'\|^2.$$ 

Combining inequality (40) with the above one, we get

$$\theta(x) - \theta(x') \geq \frac{\rho}{2} \|s - s'\|^2 - \|s\| \|JF(x') - JF(x)\| - \frac{1}{2} \|s\|^2 |D^2 F(x) - D^2 F(x')|.$$ 

Since $\theta(\cdot)$ is continuous and $F$ is twice continuously differentiable, using Corollary 5.4 and taking the limit $x' \to x$ in both sides of the above inequality, we conclude that $s' \to s$ and so $s(\cdot)$ is continuous. \hfill $\square$

5.2 The multiobjective Newton algorithm

Let us now describe the extension of the classical (scalar-valued) Newton method for the unconstrained multiobjective optimization problem (1).

Algorithm 5.6. The multiobjective Newton method (MNM)

1. Choose $\sigma \in (0, 1)$ and $x^0 \in \mathbb{R}^n$. Set $k := 0$.

2. Compute $s^k := s(x^k) = \arg\min_{s \in \mathbb{R}^n} \max_{i=1, \ldots, m} \psi_i(x^k, s)$, where $\psi_i(x^k, s) := \nabla F_i(x^k)^\top s + \frac{1}{2} \nabla^2 F_i(x^k)s$.

3. Compute $\theta(x^k) = \max_{i=1, \ldots, m} \psi_i(x^k, s^k)$. If $\theta(x^k) = 0$, then stop.

4. Choose $t_k$ as the largest $t \in \{1/2^j : j = 0, 1, 2, \ldots\}$ such that

$$F_i(x^k + ts^k) \leq F_i(x^k) + \sigma \theta(x^k) \quad \text{for all } i = 1, \ldots, m.$$ 

5. Set $x^{k+1} := x^k + t_k s^k$, $k := k + 1$ and go to Step 2.

As usual, the stopping rule at Step 3 can be replaced by $s^k = 0$ (Proposition 5.2). Now we show that if, at Step 3, $\theta(x^k) \neq 0$, the backtracking procedure prescribed in Step 4 is well-defined, i.e., it ends up in a finite number of (inner) iterations. Indeed, it suffices to note that

$$F_i(x + ts(x)) \leq F_i(x) + \sigma t \nabla F_i(x)^\top s(x) \leq F_i(x) + \sigma \theta(x)$$ 

for any nonstationary $x$, $t > 0$ and all $i$. As $JF(x)s(x) < 0$, from Lemma 2.1 it follows that there exists $\varepsilon = \varepsilon(x) > 0$ such that the first, and therefore the second, inequality in (41) hold for any $t \in (0, \varepsilon)$. So, for $x = x^k$, we will need finitely many half reductions in order to obtain $t_k \in (0, \varepsilon_k)$, where $\varepsilon_k = \varepsilon(x^k)$. Observe also that the reducing factor $1/2$ at Step 4 can be substituted by any $\nu \in (0, 1)$.

Pesquisa Operacional, Vol. 34(3), 2014
Moreover, when \( \theta(x^k) \neq 0 \) for all \( k \), we know that we have \( \theta(x^k) < 0 \), so from Step 4 it follows that

\[
F(x^{k+1}) < F(x^k) \quad \text{for all} \quad k = 0, 1, \ldots
\]

This inequality holds for \( k = 0, \ldots, k_0 - 1 \) if the algorithm stops at iteration \( k_0 \). Finally, we point out that, by Proposition 5.2, whenever the method has finite termination, it ends up with a stationary point and therefore a Pareto optimum for \( F \) (Lemma 5.1, 1). Therefore, the relevant case for the convergence analysis is the one in which the algorithm generates infinite sequences \( \{x^k\}, \{x^\ell\} \) and \( \{t_k\} \). So, from now on, we suppose that the method does not stop.

### 5.3 Convergence analysis of the MNM

Let us begin with a technical result which will be useful in the sequel.

**Proposition 5.7.** For all \( k = 0, 1, \ldots \) and all \( i = 1, \ldots, m \), we have

\[
\sum_{j=0}^{k} t_j |\theta(x^j)| \leq \sigma^{-1} (F_i(x^0) - F_i(x^{k+1})).
\]

**Proof.** From Step 3 of Algorithm 5.6, for any \( i \) and all \( j \), we have

\[
F_i(x^{j+1}) \leq F_i(x^j) + \sigma t_j |\theta(x^j)|.
\]

We complete the proof by adding up for \( j = 1, \ldots, k \) and taking into account that \( \theta(x^j) < 0 \) for all \( j \). □

We can now prove that if the sequence \( \{x^k\} \) has accumulation points, they are all Pareto optima for \( F \).

**Proposition 5.8.** If \( \bar{x} \in \mathbb{R}^n \) is an accumulation point of \( \{x^k\} \), then \( \bar{x} \) is Pareto optimum for \( F \).

**Proof.** There exists a subsequence \( \{x^{k_j}\} \) such that

\[
x^{k_j} \to \bar{x} \quad \text{as} \quad j \to \infty.
\]

Using Proposition 5.7 with \( k = k_j \) and taking the limit \( j \to \infty \), from the continuity of \( F \), we conclude that

\[
\sum_{j=0}^{\infty} t_j |\theta(x^j)| < \infty.
\]

Therefore, \( \lim_{j \to \infty} t_j \theta(x^j) = 0 \) and, in particular,

\[
\lim_{j \to \infty} t_{k_j} \theta(x^{k_j}) = 0.
\]

(42)

Suppose that \( \bar{x} \) is nonstationary, which, by Proposition 5.2, is equivalent to

\[
\theta(\bar{x}) < 0 \quad \text{and} \quad \bar{s}(\bar{x}) \neq 0.
\]

(43)
Define
\[ g: \mathbb{R}^m \to \mathbb{R}, \quad g(u) = \max_{i=1,\ldots,m} u_i. \]

Using the definition of \( \theta(\cdot) \) and Lemma 2.1, we conclude that there exists a nonnegative integer \( q \) such that
\[ g\left(F(\bar{x} + 2^{-q}\bar{s}) - F(\bar{x})\right) < \sigma 2^{-q}\theta(\bar{x}). \]

Since \( s(\cdot), \theta(\cdot) \) and \( g(\cdot) \) are continuous,
\[ \lim_{j \to \infty} s_k^j = \bar{s}, \quad \lim_{j \to \infty} \theta(x_k^j) = \theta(\bar{x}) < 0 \] (44)
and, for \( j \) large enough
\[ g\left(F(x_k^j + 2^{-q}s_k^j) - F(x_k^j)\right) < \sigma 2^{-q}\theta(x_k^j), \]
which, in view of the definition of \( g \) and Step 4 of the algorithm, implies that \( n_j \geq 2^{-q} > 0 \) for \( j \) large enough. Hence, from the second limit in (44), we conclude that \( \lim \inf_{j \to \infty} n_j \theta(x_k^j) = 0 \), in contradiction with (42). Therefore, (43) is not true and, by Proposition 5.2, \( \bar{x} \) is stationary. Whence, in view of Proposition 5.1, \( \bar{x} \) is a Pareto optimal point for \( F \).

We can now show that, in order to have full convergence to a Pareto optimum, it suffices to take the initial point \( x^0 \) in a bounded level set of \( F \).

**Theorem 5.9.** Let \( x^0 \in \mathbb{R}^n \) be in a compact level set of \( F \) and \( \{x^k\} \) a sequence generated by Algorithm 5.6. Then \( \{x^k\} \) converges to a Pareto optimum \( x^* \in \mathbb{R}^n \).

**Proof.** Let \( \Gamma_0 \) be the \( F(x^0) \)-level set of \( F \), that is, \( \Gamma_0 := \{x \in \mathbb{R}^n : F(x) \leq F(x^0)\} \). We know that the sequence \( \{F(x^k)\} \) is \( \mathbb{R}^m_+ \)-decreasing, so we have \( x_k \in \Gamma_0 \) for all \( k \). Therefore \( \{x^k\} \) is bounded, all its accumulation points belong to \( \Gamma_0 \) and so, using Proposition 5.8, we conclude that they are all Pareto optima for \( F \). Moreover, applying Lemma 3.8, 1 with \( x(x^k) = x_k \) for all \( k \), we see that all these accumulation points have the same objective value. As \( F \) is strongly (an therefore strictly) \( \mathbb{R}^m_+ \)-convex, there exists just one accumulation point of \( \{x^k\} \), say \( x^* \), and the proof is complete. □

### 5.4 Convergence rate of the MNM

Here we analyze the convergence rate of any infinite sequence \( \{x^k\} \) generated by Algorithm 5.6. First, we provide a bound for \( \theta(x^{k+1}) \) based on data at the former iterate \( x^k \). Then we show that for \( k \) large enough, full Newton steps are performed, that is to say, \( n_k = 1 \). Using this result, we prove superlinear convergence. Finally, under additional regularity assumptions, we prove quadratic convergence.

**Lemma 5.10.** For any \( k \), there exist \( \alpha_k^i \in \mathbb{R}^m \) such that \( \sum_{i=1}^m \alpha_k^i = 1, \alpha_k^i \geq 0 \) for all \( i \) with
\[ \sum_{i=1}^m \alpha_k^i \left[ \nabla F_i(x^k) + \nabla^2 F_i(x^k)x^k \right] = 0 \] (45)
and
\[ \theta(x^{k+1}) \geq -\frac{1}{2\rho} \| \sum_{i=1}^{m} a_i^k \nabla F_i(x^{k+1}) \|^2, \]  
(46)

where \( \rho > 0 \) is given by Lemma 5.1.

**Proof.** By virtue of (37)-(38) with \( x = x^k, s = s^k \), the Lagrangian condition (45) holds for any \( k \) with a vector of KKT multipliers \( \alpha = \alpha(x^k) =: \alpha^k \geq 0 \) such that \( \sum_{i=1}^{m} a_i^k = 1 \).

In order to prove (46), observe that, from Steps 2-3 of Algorithm 5.6 and the properties of \( \alpha^k \),

\[ \theta(x^{k+1}) = \sum_{i=1}^{m} a_i^k \max_j \psi_j(x^{k+1}, s^{k+1}) \]
\[ \geq \sum_{i=1}^{m} a_i^k \psi_i(x^{k+1}, s^{k+1}) \]
\[ \geq \min_j \sum_{i=1}^{m} a_i^k \psi_i(x^{k+1}, s) \]
\[ \geq \min_j \sum_{i=1}^{m} a_i^k [\nabla F_i(x^{k+1})^\top s + \frac{\rho}{2} \| s \|^2], \]

where the last inequality holds in view of Lemma 5.1 for \( x = x^{k+1} \). The function

\[ s \mapsto \sum_{i=1}^{m} a_i^k [\nabla F_i(x^{k+1})^\top s + \frac{\rho}{2} \| s \|^2] \]

is convex, so its minimal value is attained where its gradient vanishes, i.e., at

\[ s = -\frac{1}{\rho} \sum_{i=1}^{m} a_i^k \nabla F_i(x^{k+1}). \]

So

\[ \min_j \sum_{i=1}^{m} a_i^k [\nabla F_i(x^{k+1})^\top s + \frac{\rho}{2} \| s \|^2] = -\frac{1}{2\rho} \| \sum_{i=1}^{m} a_i^k \nabla F_i(x^{k+1}) \|^2. \]

The proof follows from this equality and (47). \( \square \)

Let us now see that, for any sequence defined by Algorithm 5.6, full Newton steps will be eventually taken at Step 4.

**Theorem 5.11.** Let \( x^0 \in \mathbb{R}^n \) be in a compact level set of \( F \) and \( \{x^k\} \) a sequence generated by Algorithm 5.6. Then \( \{x^k\} \) converges to a Pareto optimum point \( x^* \in \mathbb{R}^n \). Moreover, \( t_k = 1 \) for \( k \) large enough and the convergence of \( \{x^k\} \) to \( x^* \) is superlinear.
Proof. Convergence of \( \{x^k\} \) to a Pareto optimum \( x^* \) was established in Theorem 5.9. By Lemma 3.5 and Proposition 5.2, 3, we know that \( x^* \) is stationary and

\[
s(x^*) = 0. \tag{48}
\]

Since \( F \) is twice continuously differentiable, for any \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that

\[
\|\nabla^2 F_i(x) - \nabla^2 F_i(y)\| < \varepsilon \quad \text{for all } x, y \in B(x^*, \delta) \text{ and } i = 1, \ldots, m, \tag{49}
\]

where \( B(x^*, \delta) \) stands for the open ball with center in \( x^* \) and radius \( \delta \). As \( \{x^k\} \) converges to \( x^* \), using (48) and the continuity of \( s(\cdot) \) (Proposition 5.5), we conclude that \( \{s^k\} \) converges to 0. Therefore there exists \( k_\varepsilon \) such that for \( k \geq k_\varepsilon \) we have

\[
x^k, x^k + s^k \in B(x^*, \delta)
\]

and using (49) for estimating the integral remainder on Taylor’s second order expansion of \( F_i \) at \( x^k \), we conclude that for any \( i \)

\[
F_i(x^k + s^k) \leq F_i(x^k) + \nabla F_i(x^k)^\top s^k + \frac{1}{2}(s^k)^\top \nabla^2 F_i(x^k)s^k + \frac{\varepsilon}{2}\|s^k\|^2.
\]

So, using (33), (35), Corollary 5.4 and the fact that \( \sigma < 1 \), we obtain

\[
F_i(x^k + s^k) - F_i(x^k) \leq \psi_i(x^k, s^k) + \frac{\varepsilon}{2}\|s^k\|^2
\]

\[
\leq \theta(x^k) + \frac{\varepsilon}{2}\|s^k\|^2
\]

\[
= \sigma\theta(x^k) + (1 - \sigma)\theta(x^k) + \frac{\varepsilon}{2}\|s^k\|^2
\]

\[
\leq \sigma\theta(x^k) + [\varepsilon - (1 - \sigma)\rho]\|s^k\|^2 \quad \text{for all } k \geq k_\varepsilon.
\]

The above inequality shows that if \( \varepsilon < (1 - \sigma)\rho \), then, at Step 4, \( t_k = 1 \) is accepted for \( k \geq k_\varepsilon \). Suppose that

\[
\varepsilon < (1 - \sigma)\rho.
\]

Using (45) and (49) for estimating the integral remainder on Taylor’s first order expansion of \( \sum_{i=1}^m \alpha_i^k \nabla F_i \) at \( x^k \), in \( x^k + s^k \), we conclude that for any \( k \geq k_\varepsilon \) it holds

\[
\left\| \sum_{i=1}^m \alpha_i^k \nabla F_i(x^{k+1}) \right\| = \left\| \sum_{i=1}^m \alpha_i^k \nabla F_i(x^k + s^k) \right\|
\]

\[
= \left\| \sum_{i=1}^m \alpha_i^k \nabla F_i(x^k + s^k) - \sum_{i=1}^m \alpha_i^k \left[ \nabla F_i(x^k) + \nabla^2 F_i(x^k)s^k \right] \right\|
\]

\[
\leq \varepsilon \|s^k\|
\]

which, combined with (46), shows that

\[
|\theta(x^{k+1})| \leq \frac{1}{2\rho} \left[ \varepsilon \|s^k\| \right]^2.
\]
Using the above inequality and Corollary 5.4, we conclude that for $k \geq k_\varepsilon$ we have
\[
\|x^{k+1} - x^k\| \leq \frac{\varepsilon}{\rho} \|x^k - x^{k+1}\|.
\]
Therefore, if $k \geq k_\varepsilon$ and $j \geq 1$ then
\[
\|x^{k+j} - x^k\| \leq \left(\frac{\varepsilon}{\rho}\right)^j \|x^k - x^{k+1}\|. \tag{50}
\]
To prove superlinear convergence, take $0 < \xi < 1$ and define
\[
\varepsilon^* = \min\left\{1 - \sigma, \frac{\xi}{1 + 2\xi}\right\} \rho.
\]
If $\varepsilon < \varepsilon^*$ and $k \geq k_\varepsilon$, using (50) and the convergence of $\{x_k\}$ to $x^*$, we get
\[
\|x^* - x^{k+1}\| \leq \sum_{j=1}^{\infty} \|x^{k+j} - x^{k+j+1}\|
\leq \sum_{j=1}^{\infty} \left(\frac{\varepsilon}{1 + 2\xi}\right)^j \|x^k - x^{k+1}\| = \frac{\varepsilon}{1 + \xi} \|x^k - x^{k+1}\|.
\]
Hence,
\[
\|x^* - x^k\| \geq \|x^k - x^{k+1}\| - \|x^{k+1} - x^*\| \geq \frac{1}{1 + \xi} \|x^k - x^{k+1}\|.
\]
Combining the above two inequalities, we conclude that, if $\varepsilon < \varepsilon^*$ and $k \geq k_\varepsilon$, then
\[
\frac{\|x^* - x^{k+1}\|}{\|x^* - x^k\|} \leq \xi.
\]
Since $\xi$ was arbitrary in $(0, 1)$, the above quotient tends to zero and the proof is complete. \qed

Recall that in the classical Newton method for minimizing a scalar convex twice continuously differentiable function, the proof of quadratic convergence requires Lipschitz continuity of the objective function’s second derivative. Likewise, under the assumption of Lipschitz continuity of $D^2 F$, we can see that the MNM also converges quadratically.

**Theorem 5.12.** Let $x^0 \in \mathbb{R}^n$ be in a compact level set of $F$, $\{x^k\}$ a sequence generated by Algorithm 5.6 and $D^2 F$ Lipschitz continuous on $\mathbb{R}^n$. Then $\{x^k\}$ converges quadratically to a Pareto optimal solution $x^* \in \mathbb{R}^n$.

**Proof.** By Theorem 5.11, $\{x^k\}$ converges superlinearly to a Pareto optimal point $x^*$ and $t_k = 1$ for $k$ large enough. Let $\tilde{L}$ be the Lipschitz constant for $D^2 F$. Then, for $a^i \in \mathbb{R}^m$ as in Lemma 5.10, we have that $\sum_{i=1}^m a^i \nabla^2 F_i$ is also $\tilde{L}$-Lipschitz continuous. Therefore, for $k$ large enough,
\[
t_k = 1, \quad x^{k+1} - x^k = x^k \quad \text{and} \quad \left| \sum_{i=1}^m a^i \nabla^2 F_i (x^{k+1}) \right| \leq \frac{\tilde{L}}{2} \|x^k\|^2, \tag{51}
\]
where the last inequality follows from the second equality, (45) and Taylor’s development of $\sum_{i=1}^{m} q_i \nabla F_i$ at $x^k$, in $x^k + s^k$. Using the above inequality and (46), we obtain

$$-\theta(x^{k+1}) \leq \frac{1}{2\rho} \left[ \frac{\hat{L}}{2} \| s^k \|^2 \right]^2,$$

which, combined with Corollary 5.4 yields

$$\| s^{k+1} \| \leq \hat{L} \frac{\| s^k \|^2}{2\rho} \quad \text{for } k \text{ large enough.} \quad (52)$$

Take $\xi \in (0, 1)$. Since $\{x^k\}$ converges superlinearly to $x^*$, there exists $k_0$ such that for $k \geq k_0$ we have (51), (52) and

$$\| x^* - x^{k+1} \| \leq \xi \| x^* - x^k \|.$$

Therefore, applying the triangle inequality to $\| x^* - x^\ell \|$ and then to $\| x^{\ell+1} - x^\ell \|$, for $\ell \geq k_0$ we obtain

$$(1 - \xi) \| x^* - x^\ell \| \leq \| x^\ell - x^{\ell+1} \| \leq (1 + \xi) \| x^* - x^\ell \|. \quad (53)$$

Whence, using the above rightmost inequality for $\ell = k \geq k_0$ and the second equality in (51), we have

$$\| s^k \| \leq (1 + \xi) \| x^* - x^k \|,$$

while using the first inequality in (53) for $\ell = k + 1$ and the second equality in (51) (for $k + 1$ instead of $k$), we obtain

$$(1 - \xi) \| x^* - x^{k+1} \| \leq \| s^{k+1} \|.$$

Quadratic convergence of $\{x^k\}$ to $x^*$ follows from the two inequalities above and (52). \qed

### 6 FINAL REMARKS

Let us make some final comments. First, it is worth to notice that, by implementing the vector optimization versions of the steepest descent method [28] and the projected gradient method [21, 22, 25], we can guarantee convergence to Pareto optima instead of just weak Pareto points. Indeed, it suffices to take an underlying ordering (closed convex) cone $\hat{K}$ with $\mathbb{R}_+^m \setminus \{0\}$ contained in the interior of $\hat{K}$ and such that a $\hat{K}$-version of Assumption 3.6 (Assumption 4.6) in the unconstrained (constrained) case is satisfied [21, Section 6].

Without using this technique, in some cases, convergence of these two methods to (strong) Pareto optima can also be obtained. For instance, if the objective function is strictly $\mathbb{R}_+^m$-convex, then, under the assumptions of the convergence theorems, this will actually happen. Indeed, as we saw in the proof of Theorem 3.11 (Theorem 4.7), sequences generated by the MSDM (MPGM) converge to stationary points, which, in this case, are Pareto optima (Lemma 5.1, 1).

We also point out that, with these (and other) descent vector-valued methods, we are just concerned with finding a single optimum and not all the points in the optimal set. Nevertheless, from

Pesquisa Operacional, Vol. 34(3), 2014
a numerical point of view, we can expect to somehow approximate the solution set by just performing any of these methods for different initial points (e.g., [18, Section 8]). In the so-called weighting method, this kind of idea also appears. To be more precise, in such a case, one weighting vector possibly gives rise to a Pareto optimal point, which means that we should perform the method for different weights in order to find the Pareto frontier. However, in some cases, as shown in [18, Section 7], almost any choices of the weighting vectors may lead to unbounded problems. With arbitrary choices of initial points, we do not have this disadvantage in the descent vector-valued methods discussed here. We believe this is one of the main motivations for the study of such methods.

ACKNOWLEDGMENTS

This work was supported by FAPESP (grant 2010/20572-0) and by FAPERJ/CNPq through PRONEX-Optimization. We are also grateful to Benar F. Svaiter for his suggestions and to an anonymous referee for his helpful comments.

REFERENCES


