Math Meth Oper Res (2000) 51:479–494

Mathematical Methods
of Operations Research

Steepest descent methods for multicriteria optimization

Jörg Fliege[1], **Benar Fux Svaiter**[2]

[1] Fachbereich Mathematik, Universität Dortmund, 44221 Dortmund, Germany
(e-mail: fliege@math.uni-dortmund.de)
[2] Instituto de Matemática Pura e Aplicada, Jardim Botanico, Estrada Dona Castorina 110,
Rio de Janeiro, RJ 22460-320, Brasil (e-mail: benar@impa.br)

Abstract. We propose a steepest descent method for unconstrained multicriteria optimization and a "feasible descent direction" method for the constrained case. In the unconstrained case, the objective functions are assumed to be continuously differentiable. In the constrained case, objective and constraint functions are assumed to be Lipshitz-continuously differentiable and a constraint qualification is assumed. Under these conditions, it is shown that these methods converge to a point satisfying certain first-order necessary conditions for Pareto optimality. Both methods do not scalarize the original vector optimization problem. Neither ordering information nor weighting factors for the different objective functions are assumed to be known. In the single objective case, we retrieve the Steepest descent method and Zoutendijk's method of feasible directions, respectively.

Key words: Multicriteria optimization, multi-objective programming, vector optimization, Pareto points, steepest descent

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, 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 the same or smaller 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 engi-

Manuscript received: July 1999/final version received: November 1999

neering design [6] (especially truss optimization [5]) location science [2], statistics [1], management science [7] (especially portfolio analysis [14]), etc.

The main solution strategy for multicriteria optimization problems is the scalarization approach due to Geoffrion [8]. Here, one or several parameterized single-objective (i. e. classical) optimization problems are solved. The disadvantage of this approach is that the choice of the parameters is not known in advance, leaving the modeler and the decision-maker with the burden of choosing them. Furthermore, this method computes only *proper* Pareto-optimal points. Other scalarization techniques are parameter-free [4, 5], but try to compute a discrete approximation to the whole set of Pareto-optimal points.

Parameter-free multicriteria optimization techniques usually use 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 [12], adding an additional burden to the task of the decision maker.

In this paper, we propose two parameter-free optimization methods for computing a point satisfying first-order necessary conditions for multicriteria optimization. Neither ordering information nor weighting factors for the different objective functions are assumed to be known.

2 Notation

Denote by $\mathbb{R}$ the set of real numbers, by $\mathbb{R}_+$ the set of nonnegative real numbers and by $\mathbb{R}_{++}$ the set of strictly positive real numbers. The Euclidean norm in $\mathbb{R}^n$ will be denoted by $\|\cdot\|$. The problem is to find a *Pareto optimum* of

$$F : \mathbb{R}^n \to \mathbb{R}^m,$$

i. e. a point $z \in \mathbb{R}^n$ such that there does not exist a point $y \in \mathbb{R}^n$ with $F(y) \leq F(z)$ and $F(y) \neq F(z)$. Here, the inequality sign $\leq$ between vectors is to be understood in a componentwise sense. Likewise, in what follows, a strict inequality $F(y) < F(z)$ is to be understood componentwise, too. A point $z \in \mathbb{R}^n$ is called *locally Pareto optimal*, if there is a neighborhood $U \subseteq \mathbb{R}^n$ of z such that there does not exist a point $y \in U$ with $F(y) \leq F(z)$ and $F(y) \neq F(z)$. Note that each local Pareto optimal point is globally Pareto optimal as soon as all functions F_i $(i = 1, \ldots, m)$ are convex. In what follows, we assume that the function F is continuously differentiable. For $x \in \mathbb{R}^n$ the Jacobian of F at x is denoted by $JF(x)$, an $m \times n$ matrix with entries

$$(JF(x))_{i,j} = \frac{\partial F_i}{\partial x_j}(x).$$

Let range(A) denotes the range of the linear mapping given by the matrix A. A necessary condition for a point $x \in \mathbb{R}^n$ to be locally Pareto optimal is

$$\text{range}(JF(x)) \cap (-\mathbb{R}_{++})^m = \emptyset, \tag{1}$$

see, e.g., [10, 9].

Therefore we will call points that satisfy the above condition (1) *Pareto critical points*. If a point x is not Pareto critical then there exists a direction $v \in \mathbb{R}^n$ satisfying

$$JF(x)v \in (-\mathbb{R}_{++})^m,$$

i.e. a descent direction for the vector-valued objective function F. The idea of the general algorithm is now straightforward: choose an x and check if (1) holds. If not, compute a direction v and make a step with a suitably chosen steplength from x along v. This results in a new point, and the scheme can be repeated.

3 Computing the search direction

Suppose that we have given a point $x \in \mathbb{R}^n$. Define

$$A = JF(x)$$

and the function $f_x : \mathbb{R}^n \to \mathbb{R}^m$ by

$$f_x(v) := \max\{(Av)_i \,|\, i = 1, \ldots, m\}.$$

Of course, f_x is convex (as the maximum of linear functions) and positive homogeneous:

$$f_x(\lambda v) = \lambda f_x(v) \quad \forall \lambda \in \mathbb{R}_+.$$

Consider the unconstrained minimization problem

$$\begin{aligned} \min \quad & f_x(v) + \frac{1}{2}\|v\|^2 \\ \text{subject to} \quad & v \in \mathbb{R}^n. \end{aligned} \tag{2}$$

Since the objective function is proper, closed, and strongly convex, it has always a (unique) solution. Note that a simple reformulation to get rid of the nondifferentiabilities would be

$$\begin{aligned} \min \quad & \alpha + \frac{1}{2}\|v\|^2 \\ \text{subject to} \quad & (Av)_i \le \alpha, \quad i = 1, \ldots, m, \end{aligned} \tag{3}$$

which is a convex quadratic problem with linear inequality constraints.

Lemma 1. *Let $v(x)$ and $\alpha(x)$ be the solution and the optimum value of problem (2), respectively.*

1. *If x is Pareto critical, then $v(x) = 0 \in \mathbb{R}^n$ and $\alpha(x) = 0$.*
2. *If x is not Pareto critical, then $\alpha(x) < 0$ and*

$$f_x(v(x)) \leq -(1/2)\|v(x)\|^2 < 0,$$

$$(JF(x)v(x))_i \leq f_x(v(x)), \quad i = 1, \ldots, m.$$

3. *The mappings* $x \mapsto v(x)$ *and* $x \mapsto \alpha(x)$ *are continuous.*

Proof: All claims are immediately clear from the definition of problem (2). □

Remark 1. If $m = 1$, we fall back to single-objective minimization. In this case, $f_x(v) = (\nabla F(x))^\top v$ and

$$v(x) = -\nabla F(x).$$

So, we retrieve the steepest descent direction in the case $m = 1$.

Instead of solving problem (2) *exactly*, it is interesting for algorithmic purposes to deal with *inexact* solutions. So, if x is not Pareto critical, we say that v is an approximative solution of (2) with tolerance $\sigma \in]0, 1]$ if

$$f_x(v) + \frac{1}{2}\|v\|^2 \leq \sigma\alpha(x), \tag{4}$$

where, as above, $f_x(v) := \max_i(JF(x)v)_i$ and $\alpha(x)$ is the optimum value of problem (2). Observe that for $\sigma = 1$ only the exact solution satisfies the above inequality.

Define, for $A \in \mathbb{R}^{m\times n}$,

$$\|A\|_{\infty,2} := \max_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|}.$$

Then $\|\cdot\|_{\infty,2}$ is a norm in $\mathbb{R}^{m\times n}$. It is easy to prove that

$$\|A\|_{\infty,2} := \max_{i=1,\ldots,m} \|A_{i,\cdot}\|$$

$$= \max_{i=1,\ldots,m} \left(\sum_{j=1}^n A_{i,j}^2\right)^{1/2}.$$

Lemma 2. *Suppose* x *is not Pareto critical and that* v *is an approximated solution of (2) with tolerance* $\sigma \in]0, 1]$. *Then*

$$\|v\| \leq 2\|A\|_{\infty,2}.$$

3.1 Other possibilities for the search direction

Of course, there is no need for the specific choice of $(1/2)\|\cdot\|^2$ as the function which is added to f_x to define problem (2). In fact, any proper closed strictly convex function $g : \mathbb{R}^n \to \mathbb{R}$ with $g(0) = 0$, $g(v) > 0$ $(v \neq 0)$ and $g(v) = o(\|v\|)$ $(v \to 0)$ can be used to define the minimization problem $\min_{v \in \mathbb{R}^n} f_x(v) + g(v)$. The results of Lemma 1 carry over to this generalized case.

Now let us discuss another possibility for choosing descent search directions at x. Instead of solving (2) we could take v as a solution of

$$\begin{aligned} \min \quad & f_x(v) \\ \text{subject to} \quad & \|v\|_\infty \leq 1. \end{aligned} \tag{5}$$

This problem can also be reformulated as

$$\begin{aligned} \min \quad & \beta \\ \text{subject to} \quad & (Av)_i \leq \beta, \quad i = 1, \ldots, m, \\ & \|v\|_\infty \leq 1, \end{aligned} \tag{6}$$

which is a linear problem. Let $\beta(x)$ be the optimal function value of (6). We have $\beta(x) \leq 0$, because $(0, 0) \in \mathbb{R}^{n+1}$ is feasible for (6). As above, if $\beta(x) = 0$, the point x is Pareto critical. If $\beta(x) < 0$, it is possible to compute a descent direction by solving (6).

Lemma 3. *Let $V(x)$, $\beta(x)$ be the solution-set and the optimum value of problem (5), respectively.*

1. *If x is Pareto critical, then $0 \in V(x)$ and $\beta(x) = 0$.*
2. *If x is not Pareto critical, then $\beta(x) < 0$ and for any $v \in V(x)$ we have*

$$(JF(x)v)_i \leq f_x(v) < 0, \quad i = 1, \ldots, n.$$

3. *The mapping $x \mapsto \beta(x)$ is continuous.*
4. *If $x^{(k)}$ converges to $\bar{x}$, $v^{(k)} \in V(x^{(k)})$ and $v^{(k)}$ converges to $\bar{v}$ then $\bar{v} \in V(\bar{x})$.*

The choice of the constraint $\|v\|_\infty \leq 1$ can be replaced by the more general constraint $v \in D$ where $D \subset \mathbb{R}^n$ is a closed compact set with $0 \in \operatorname{int}(D)$. The obvious advantage of the choice made is, of course, the simple linear structure of problem (6).

No matter which auxiliary problem is used to compute a descent direction, note that there is no need to solve these auxiliary problems exactly. A descent direction is found as soon as we have a function value in the auxiliary problem which satisfies (4).

4 Computing the step length

Suppose that we have a direction $v \in \mathbb{R}^n$ with $JF(x)v < 0$. To compute a steplength $t > 0$ we use an Armijo-like rule. Let $\beta \in \,]0, 1[$ be a prespecified constant. The condition to accept t is

$$F(x+tv) \le F(x) + \beta t JF(x)v. \tag{7}$$

We start with $t = 1$ and while this condition is not satisfied, we set

$$t := t/2.$$

Finiteness of this procedure follows from the fact that (7) holds strictly for $t > 0$ small enough.

Lemma 4. *If F is differentiable and $JF(x)v < 0$, then there exist some $\varepsilon > 0$ (which may depend on x, v and β) such that*

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

for any $t \in]0, \varepsilon]$.

Proof: Since we assumed F to be differentiable, we have

$$F(x+h) = F(x) + JF(x)h + R(h)$$

with

$$\lim_{h \to 0} \frac{|R_i(h)|}{\|h\|} = 0, \quad i = 1, \dots, m.$$

Define

$$a := \max_i (JF(x)v)_i.$$

Observe that $a < 0$ and $v \neq 0$. Since $\beta < 1$, there exists an $\varepsilon > 0$ such that

$$0 < t \le \varepsilon \Rightarrow \frac{|R(tv)_i|}{\|tv\|} < \frac{(1-\beta)|a|}{\|v\|}, \quad i = 1, \dots, m.$$

Hence, for $0 < t \le \varepsilon$ we get

$$|R_i(tv)| < t(1-\beta)|a|, \quad i = 1, \dots, m,$$

and, since $|a| = -a = \min_i -(JF(x)v)_i$,

$$R(tv) < -t(1-\beta)JF(x)v.$$

Therefore,

$$\begin{aligned} F(x+tv) &= F(x) + tJF(x)v + R(tv) \\ &< F(x) + tJF(x)v - t(1-\beta)JF(x)v \\ &= F(x) + t\beta JF(x)v \end{aligned}$$

for $0 < t \le \varepsilon$. $\square$

Remark 2. Let $|||\cdot|||$ be a norm in $\mathbb{R}^n$ and $|||\cdot|||^*$ be the subordinate dual norm. Suppose that JF is Lipshitz-continuous and

$$|||\nabla F_i(y) - \nabla F_i(\tilde{y})|||^* \leq L|||y - \tilde{y}|||, \quad i = 1, \ldots, m,$$

for all $y, \tilde{y} \in \mathbb{R}^n$.

If $JF(x)v < 0$, define

$$\tilde{\varepsilon} := -\frac{2(1-\beta)}{L} \frac{\max_i (JF(x)v)_i}{(|||v|||)^2} > 0.$$

Then

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

for any $t \in [0, \tilde{\varepsilon}]$.

In the conditions of the above remark, if $|||\cdot|||$ is the euclidian norm and $v = v(x)$, then $\tilde{\varepsilon} \geq (1-\beta)/L$.

5 The complete algorithm

Algorithm 1.

1. Choose $\beta \in]0, 1[$, $\sigma \in]0, 1]$, and $x^{(1)} \in \mathbb{R}^n$, set $k := 1$.
2. If $x^{(k)}$ is Pareto critical, stop.
3. Compute $v^{(k)}$, an approximative solution of (2) at $x = x^{(k)}$ with tolerance σ.
4. Compute a steplength $t_k \in]0, 1]$ as the maximum of

$$T_k := \{t = 1/2^j \mid j \in \mathbb{N}, F(x^{(k)} + tv^{(k)}) \leq F(x^{(k)}) + \beta t JF(x^{(k)})v^{(k)}\}.$$

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

Observe that if Step 3 is reached at iteration k, then $x^{(k)}$ is *not* Pareto critical and

$$\max_i (JF(x^{(k)})v^{(k)})_i \leq \max_i (JF(x^{(k)})v^{(k)})_i + \frac{1}{2}\|v^{(k)}\|^2 \leq \sigma\alpha(x^{(k)}) < 0.$$

So, $JF(x^{(k)})v^{(k)} < 0$. Hence at this iteration T_k is not empty and Step 4 is well defined. Indeed, note that the computation of the steplength in Step 4 can be done by the Armijo-like rule outlined above.

6 Convergence of the steepest descent method

Observe that if Algorithm 1 terminates after a finite number of iterations, it terminates at a Pareto critical point. From now on we suppose that an infinite sequence is generated, so $\alpha(x^{(k)}) \neq 0$ for all $k \in \mathbb{N}$.

We present a simple application of the standard convergence argument for the steepest descent method.

Theorem 1. *Every accumulation point of the sequence $(x^{(k)})_k$ produced by Algorithm 1 is a Pareto critical point. If the function F has bounded level sets in the sense that the set $\{x \in \mathbb{R}^n \mid F(x) \leq F(x^{(1)})\}$ is bounded, then the sequence $(x^{(k)})_k$ stays bounded and has at least one accumulation point.*

Proof. Let y be an accumulation point of the sequence $(x^{(k)})_k$ and let $v(y)$ and $\alpha(y)$ be the solution and the optimum value of (2) at y:

$$v(y) := \arg\min(f_y(v) + (1/2)\|v\|^2), \quad \alpha(y) := \min_v(f_y(v) + (1/2)\|v\|^2),$$

where

$$f_y(v) := \max_i(JF(y)v)_i.$$

According to Lemma 1 it is enough to prove that $\alpha(y) = 0$.

Clearly, the sequence $(F(x^{(k)}))_k$ is componentwise strictly decreasing and we have

$$\lim_{k\to\infty} F(x^{(k)}) = F(y).$$

Therefore,

$$\lim_{k\to\infty} \|F(x^{(k)}) - F(x^{(k+1)})\| = 0.$$

But

$$F(x^{(k)}) - F(x^{(k+1)}) \geq -t_k\beta JF(x^{(k)})v^{(k)} \geq 0,$$

and therefore

$$\lim_{k\to\infty} t_k JF(x^{(k)})v^{(k)} = 0. \tag{8}$$

Observe that $t_k \in]0, 1]$ for all k. Now take a subsequence $(x^{(k_u)})_u$ converging to y. We will consider two possibilities:

$$\limsup_{u\to\infty} t_{k_u} > 0$$

and

$$\limsup_{u\to\infty} t_{k_u} = 0.$$

First case. In this case, there exist a subsequence $(x^{(k_\ell)})_\ell$ converging to y and satisfying

$$\lim_{\ell \to \infty} t_{k_\ell} = \bar{t} > 0$$

Using (8) we conclude that

$$\lim_{\ell \to \infty} JF(x^{(k_\ell)})v^{(k_\ell)} = 0,$$

which also implies

$$\lim_{\ell \to \infty} \alpha(x^{(k_\ell)}) = 0.$$

Since $x \mapsto \alpha(x)$ is continuous, we conclude that $\alpha(y) = 0$, so y is Pareto critical.

Second case. Due to the results in Lemma 2 we know that the sequence $(v^{(k_u)})_u$ is bounded. Therefore we can take a subsequence $(x^{(k_r)})_r$ of $(x^{(k_u)})_u$ such that the sequence $(v^{(k_r)})_r$ also converges to some $\bar{v}$. Note that for all r we have

$$\max_i (JF(x^{(k_r)})v^{(k_r)})_i \leq \sigma\alpha(x^{(k_r)}) < 0.$$

Therefore, passing onto the limit $r \to \infty$ we get

$$\frac{1}{\sigma} \max_i (JF(y)\bar{v})_i \leq \alpha(y) \leq 0. \tag{9}$$

Take some $q \in \mathbb{N}$. For r large enough,

$$t_{k_r} < \frac{1}{2^q},$$

which means that the Armijo condition (7) is not satisfied for $t = 1/2^q$, i.e.,

$$JF(x^{(k_r)} + (1/2^q)v^{(k_r)}) \not\leq F(x^{(k_r)}) + \beta(1/2^q)JF(x^{(k_r)})v^{(k_r)}$$

(for r large enough). Passing onto the limit $r \to \infty$ (along a suitable subsequence, if necessary) we get

$$F_j(y + (1/2^q)\bar{v}) \geq F_j(y) + \beta(1/2^q)(JF(y)\bar{v})_j$$

for at least one $j \in \{1, \ldots, m\}$. Note that this inequality holds for any $q \in \mathbb{N}$. From Lemma 4 it follows that

$$\max_i (JF(y)\bar{v})_i \geq 0,$$

which, together with (9), implies $\alpha(y) = 0$. So, again we conclude that y is Pareto critical. □

7 Towards an implementable method

Theorem 1 guarantees that under suitable conditions the accumulation points of the sequence $(x^{(k)})_k$ are Pareto critical. But in practice, every algorithm should terminate in finite time. So, a stopping criterion other than the theoretical one used in Step 2 of Algorithm 1 should be used.

In single objective optimization, when the optimal value is neither known nor estimated *a priori*, the norm of the gradient of the objective function is frequently used in a stopping criterion. The algorithm in use should terminate as soon as this norm is small enough. This, however, is not possible in multicriteria optimization since the gradients of the objective functions may not vanish at a Pareto optimal/critical point.

But since the direction provided by (2) generalizes the steepest descent direction, the optimal value of the direction search problem (2) can be used as a stopping criterion in an implementation of the algorithm. In this case one would replace Steps 2 and 3 of the algorithm by

2. Compute $v^{(k)}$, an approximative solution of (2) at $x = x^{(k)}$ with tolerance σ. Let α_k be the function value of (2) at $v^{(k)}$.
3. If $-\tau < \alpha_k$, stop.

Here, $\tau > 0$ is a prespecified accuracy value. As it can be seen, we do not need to compute the exact optimal function value of the direction search program, but only an upper bound for it. This is exactly what a generic minimization technique employed to solve (2) provides.

Nevertheless, the knowledge of the optimal value $\alpha(x^{(k)})$ is useful to accept a direction v as an approximated solution. This can be seen in inequality (4), where the exact optimal value explicitly appears. The use of this inequality in an iterative algorithm solving the direction search program makes it necessary to compute good bounds for $\alpha(x^{(k)})$.

Dual (and primal dual) algorithms (and this includes log barrier methods) provide lower bounds for the optimum value, which obviates the exact computation of the optimum value of problem (2). Furthermore, to work with a primal-dual reformulation of the direction search program is quite natural, because the dual of (3) is

$$
\begin{aligned}
\max \quad & -\frac{1}{2}\left\| \sum_{i=1}^{m} \theta_i A_{i,\cdot} \right\|^2 \\
\text{subject to} \quad & \sum_{i=1}^{m} \theta_i = 1, \\
& \theta \geq 0.
\end{aligned}
$$

The same considerations apply to search directions based on the linear programming formulation (6). Here, the dual is

$$\begin{aligned}
\max \quad & -\left\| \sum_{i=1}^{m} \theta_i A_{i,\cdot} \right\|_1 \\
\text{subject to} \quad & \sum_{i=1}^{m} \theta_i = 1, \\
& \theta \geq 0.
\end{aligned}$$

On the other hand, it has to be noted that iteration steps in primal-dual methods are more expensive than the corresponding steps in pure primal algorithms. At the present moment, it is therefore not clear which formulation and which method one should use to solve the direction search problem (2).

It is interesting to note that both duals to the direction search programs proposed are standard goal programming problems. Goal programming problems arise naturally when reformulating multicriteria optimization problems [13, 3], but they appear here as the duals of "linearized" nonlinear vector optimization problems. Both duals above are special cases of the general problem

$$\begin{aligned}
\min \quad & \gamma(u) \\
\text{subject to} \quad & u \in \operatorname{conv}\{A_{i,\cdot} \mid i = 1, \ldots, m\},
\end{aligned}$$

where γ is a gauge (i. e. a norm-like distance function) in the $\mathbb{R}^n$. Using the results found in [3], it is easy to see that

$$\begin{aligned}
\min \quad & \max\{-(Av)_i \mid i = 1, \ldots, m\} \\
\text{subject to} \quad & \gamma^\circ(v) \leq 1
\end{aligned}$$

is the corresponding primal problem. Clearly, both direction search programs described in Section 3 are special cases of this generalization.

8 The constrained case

Suppose that a set of feasible points $G \subseteq \mathbb{R}^n$ is given. A point z is Pareto optimal for F restricted on G if it is feasible (i. e., $z \in G$) and there does not exist a point $y \in G$ with $F(y) \leq F(z)$ and $F(y) \neq F(z)$. As in the unconstrained case, a feasible point $z \in G$ is called locally Pareto optimal, if there is a neighborhood $U \subseteq \mathbb{R}^n$ of z such that there does not exist a point $y \in U \cap G$ with $F(y) \leq F(z)$ and $F(y) \neq F(z)$.

Given $z \in G$, a vector $v \in \mathbb{R}^n$ is a *tangent direction* of G at z if there exists a sequence $(z^{(k)})_k \subseteq G$ and a scalar $\lambda \in \mathbb{R}_+$ such that

$$\lim_{k\to\infty} z^{(k)} = z \quad \text{and} \quad \lim_{k\to\infty} \lambda \frac{z^{(k)} - z}{\|z^{(k)} - z\|} = v.$$

The set of all tangent directions of G at z is called the *contingent cone* of G at z and will be denoted by $T(G, z)$.

Observe that if z is locally Pareto optimal and $v \in T(G, z)$, then $(\nabla F_i(z))^\top v < 0$ can not hold for all $i = 1, \ldots, m$. So, a *necessary* condition for local Pareto optimality is

$$JF(T(G, z)) \cap (-\mathbb{R}^m_{++}) = \emptyset, \tag{10}$$

which is equivalent to saying that the system

$$(\nabla F_i(z))^\top v < 0, \quad \forall i = 1, \ldots, m, \quad v \in T(G, z), \tag{11}$$

has no solutions. Now suppose that G is given by

$$G = \{x \in \mathbb{R}^n \mid g_j(x) \leq 0, j = 1, \ldots, \ell\}$$

with continuously differentiable functions $g_j : \mathbb{R}^n \to \mathbb{R}$. The set of *active* constraints at a feasible point z will be denoted by $I_0(z)$,

$$I_0(z) := \{j \mid g_j(z) = 0\},$$

while the *linearized cone* at z, denoted by $C(z)$ is given by

$$C(z) := \{v \in \mathbb{R}^n \mid (\nabla g_j(z))^\top v \leq 0 \ \forall j \in I_0(z)\}.$$

Suppose now that a constraint qualification holds, namely that

$$C(z) = T(G, z)$$

holds for all feasible z. Under this assumption, condition (10) (and therefore (11)) becomes equivalent to saying that the system

$$\begin{aligned} \nabla F_i(x))^\top v &< 0, \quad \forall i \\ \nabla g_j(x))^\top v &\leq 0 \quad \forall \text{ active } j \end{aligned} \tag{12}$$

has no solution $v \in \mathbb{R}^n$.

Let $\hat{A}$ be the matrix obtained by adding rows $(\nabla g_j(x))^\top$ for all active constraints j to the Jacobian $JF(x)$. Necessary for the condition above is

$$\text{range}(\hat{A}) \cap ((-\mathbb{R}_{++})^m) = \emptyset.$$

We see that we can use the same steepest descent algorithm as above, by adding some rows in each step according to the active constraints, and use a step length control to enforce feasibility as well as sufficient decrease. The auxiliary program is virtually the same as (2), the only difference being that the matrix A is replaced by the matrix $\hat{A}$.

To use the above idea in a convergent algorithm, we replace (2) by the following program. Let $I := \{1, \ldots, \ell\}$. For a given $\varepsilon \in \mathbb{R}_+$ and $x \in \mathbb{R}^n$ define the index set I_ε by

$$I_\varepsilon(x) := \{j \in I \mid g_j(x) \geq -\varepsilon\}.$$

We then use the direction search program

$$\begin{aligned} \min \quad & \alpha \\ \text{subject to} \quad & (\nabla F_i(x))^\top v \leq \alpha, \quad i = 1, \dots, m, \\ & (\nabla g_j(x))^\top v \leq \alpha, \quad j \in I_\varepsilon(x), \\ & \|v\|_\infty \leq 1, \end{aligned} \tag{13}$$

and denote the optimal function value of this problem by $\alpha(x, \varepsilon)$. A simple application of Farkas's Lemma leads to the observation that $\alpha(z, 0) = 0$ holds for a feasible point z if and only if there exist vectors $\lambda \in \mathbb{R}^m_+ \setminus \{0\}$, $\theta \in \mathbb{R}^\ell_+$ with

$$\sum_{i=1}^m \lambda_i \nabla F_i(z) + \sum_{j=1}^\ell \theta_i \nabla g_i(z) = 0$$

and

$$\sum_{j=1}^\ell \theta_i g_i(z) = 0.$$

In turn, these conditions are necessary for Pareto optimality as long as the constraint qualification made above holds [11]. Moreover, if the functions involved are convex and $\lambda \in \mathbb{R}^m_{++}$ holds, then these conditions are sufficient for Pareto optimality of z (cmp. [11]).

Following the results found in Zoutendijk [15], we have the next two lemmas.

Lemma 5. *Let $(x^{(k)})_k$ be a sequence in G with $\lim_{k\to\infty} x^{(k)} = y$ and let $(\varepsilon_k)_k$ be a sequence in $\mathbb{R}_+$ with $\lim_{k\to\infty} \varepsilon_k = 0$. Then we have*

$$\limsup_{k\to\infty} \alpha(x^{(k)}, \varepsilon_k) \leq \alpha(y, 0).$$

Lemma 6. *For all $i \in I$, let ∇g_i be Lipshitz-continuous with Lipshitz-constant L with respect to the ∞-norm. Let $x \in G$, $v \in \mathbb{R}^n$, $\varepsilon \in \mathbb{R}_+$ and $\varrho \in -\mathbb{R}_{++}$ be given such that*

$$(\nabla g_j(x))^\top v \leq \rho \quad \textit{for all } j \in I_\varepsilon(x).$$

It then follows that

$$x + tv \in G \quad \textit{for all } t \in [0, \min\{1, \varepsilon/L, -2\rho/L\}].$$

These results enable us to prove our main theorem.

Theorem 2. *Let JF and ∇g_i $(i \in I)$ be Lipshitz-continuous and let $x^{(1)} \in G$. For a fixed $\varepsilon \in \mathbb{R}_{++}$, let the sequence $(x^{(k)})_k \subset G$ be generated according to the direction program (13), where the step lengths are chosen according to the Armijo criterion (7). Then we either have*

$$\lim_{k\to\infty} F_i(x^{(k)}) = -\infty$$

for all i, or we have

$$\lim_{k\to\infty} \alpha(x^{(k)}, \varepsilon) = 0. \tag{14}$$

Proof: It is

$$F(x^{(k+1)}) \leq F(x^{(k)}) + t_k JF(x^{(k)})v^{(k)} + t_k^2 \frac{L}{2} e,$$

where $e = (1, \ldots, 1)^\top$ and L is the Lipshitz constant of JF and the g_i. Suppose that (14) does not hold. But we always have $\alpha(x^{(k)}, \varepsilon) \leq 0$ for all k, which means that there exists a number $\varrho \in -\mathbb{R}_{++}$ and a subsequence with $\alpha(x^{(k_u)}, \varepsilon) < \varrho$. Then,

$$F(x^{(k_u+1)}) \leq F(x^{(k_u)}) + t_{k_u}\varrho e + t_{k_u}^2 \frac{L}{2} e \tag{15}$$

for all u. Without loss of generality, assume $-2\varrho \leq \varepsilon$ and $-2\varrho \leq L$. According to Lemma 6, $x^{(k)} + tv^{(k)} \in G$ holds as long as $t \leq -2\varrho/L$. But Lemma 4 and the accompanying remark tell us that the Armijo criterion holds for $0 \leq t \leq -2\varrho(1-\beta)/L$, and so we have

$$t_{k_u} \in \left[-\varrho\frac{1-\beta}{L}, -2\varrho\frac{1-\beta}{L}\right].$$

A simple calculation based on (15) shows that

$$F(x^{(k_u+1)}) \leq F(x^{(k_u)}) + \varrho^2 \frac{1-\beta}{L} \max\left\{-\frac{1}{2}(1+\beta), -2\beta\right\}$$

holds. As a result, $\lim_{k\to\infty} F_i(x^{(k)}) = -\infty$ for all i. □

We can now adapt Zoutendijk's method of feasible directions to the multicriteria case considered here.

Algorithm 2.

1. Choose $\beta \in]0, 1[$, $\varepsilon_1 \in \mathbb{R}_{++}$, and $x^{(1)} \in G$, set $k := 1$.
2. Solve (13), thereby computing a direction $v^{(k)}$ and an optimal function value $\alpha(x^{(k)}, \varepsilon_k)$ for the auxiliary problem.
3. If $\alpha(x^{(k)}, \varepsilon_k) \geq -\varepsilon_k$, set $\varepsilon_{k+1} := \varepsilon_k/2$, $x^{(k+1)} := x^{(k)}$, $k := k+1$, and goto Step 2.

4. Compute a steplength $t_k \in]0, 1]$ such that

$$F(x^{(k)} + t_k v^{(k)}) \leq F(x^{(k)}) + \beta t_k JF(x^{(k)}) v^{(k)}$$

and $x^{(k)} + t_k v^{(k)} \in G$.

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

Theorem 3. *Let JF and ∇g_i $(i \in I)$ be Lipshitz-continuous and let the sequence $(x^{(k)})_k$ generated by Algorithm 2 be bounded. It then follows that the sequence $(x^{(k)})_k$ has an accumulation point y with $\alpha(y, 0) = 0$.*

Proof: The proof is identical to the convergence theorem of Zoutendijk's P1-method and is therefore omitted here. □

Note that when F has bounded level sets, the sequence $(x^{(k)})_k$ is necessarily bounded.

Similar to the unconstrained case, it is not necessary to solve (13) exactly. Instead, it is sufficient to stop as soon as a function value α has been computed with $\alpha < -\varepsilon_k$. In case $\alpha(x^{(k)}, \varepsilon_k) \geq -\varepsilon_k$ holds for the optimal function value $\alpha(x^{(k)}, \varepsilon_k)$, dual bounds can be used to stop prematurely.

9 Final remarks

9.1 Variations of the basic algorithm

All convergence properties of Algorithm 1 remain valid if we take at Step 3 the vector $v^{(k)}$ as the solution of Problem (6) at $x = x^{(k)}$. This variation has the advantage of requiring a solution of a linear problem instead of a quadratic one in each iteration. Again, approximated solutions of problem (6), defined in a similar way to (4), may be used.

9.2 Extensions

Observe that only first order information on the objective functions has been used. As a consequence, it is clear that the methods proposed in this paper should be seen more as first steps towards efficient methods, rather than as efficient methods in themselves. This can also be seen from the fact that the algorithms proposed in this paper reduce to the steepest descent method resp. Zoutendijk's method in case only one objective is given. Therefore, an interesting topic for future research is an elaboration of a Newton Method-like algorithm for multicriteria optimization.

The constrained multicriteria optimization problem is another topic for future research.

References

[1] Carrizosa E, Frenk JBG (1996) Dominating sets for convex functions with some applications. Preprint TI 96-156/9, Tinbergen Institute, The Netherlands

[2] Carrizosa E, Conde E, Munoz M, Puerto J (1995) Planar point-objective location problems with nonconvex constraints: a geometrical construction. Journal of Global Optimization 6:77–86
[3] Carrizosa E, Fliege J (1999) Generalized goal programming: Polynomial methods and applications. Report BEIF/108, Vakgroep Beleidsinformatica, Vrije Universiteit Brussel
[4] Das I, Dennis JE (1997) A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multi-criteria optimization problems. Structural Optimization 14:63–69
[5] Das I, Dennis JE (1998) Normal-boundary intersection: a new method for generating Pareto optimal points in nonlinear multicriteria optimization problems. SIAM Journal on Optimization 8:631–657
[6] Eschenauer H, Koski J, Osyczka A (1990) Multicriteria design optimization. Springer-Verlag, Berlin
[7] Evans G (1984) Overview of techniques for solving multiobjective mathematical programs. Management Science 30:1268–1282
[8] Geoffrion AM (1968) Proper efficiency and the theory of vector maximization. Journal of Mathematical Analysis and Applications 22:618–630
[9] Göpfert A, Nehse R (1990) Vektoroptimierung: Theorie, Verfahren und Anwendungen. B. G. Teubner Verlagsgesellschaft, Leipzig
[10] Luc DT (1988) Theory of vector optimization. Springer-Verlag, Berlin
[11] Marusciac I (1982) On Fritz-John-type optimality criterion in multi-objective optimization. Mathematica – Revue d'analyse numerique et de théorie de l'approximation 11:109–114
[12] Miettinena K, Mäkelä MM (1995) Interactive bundle-based method for nondifferentiable multiobjective optimization: NIMBUS. Optimization 34:231–246
[13] Romero C (1991) Handbook of critical issues in goal programming. Pergamon Press, Oxford
[14] White DJ (1998) Epsilon-dominating solutions in mean-variance portfolio analysis. European Journal of Operational Research 105:457–466
[15] Zoutendijk G (1960) Methods of feasible directions. Elsevier, Amsterdam