

A derivative-free descent method in set optimization

Johannes Jahn

Received: 13 November 2013 / Published online: 11 July 2014
© Springer Science+Business Media New York 2014

Abstract Based on a vectorization result in set optimization with respect to the set less order relation, this paper shows how to relate two nonempty sets on a computer. This result is developed for generalized convex sets and polyhedral sets in finite dimensional spaces. Using this approach a numerical method for the determination of optimal scenarios is presented. A new derivative-free descent method for the solution of set optimization problems is given together with numerical results in low dimensions.

Keywords Set optimization · Set less order relation · Vectorization

Mathematics Subject Classification 90C56 · 90C29 · 06A06

1 Introduction

Set optimization has developed from an extension of continuous optimization to problems with set-valued maps. For early contributions see the papers published by Borwein [1] in 1977 and Oettli [21] in 1980. Although most of the papers on set optimization work with the notion of a minimizer and variants of it, nowadays one works with a more realistic order relation for the comparison of sets which has been introduced to optimization by Kuroiwa (e.g., see [12]; a first publication has been given by Kuroiwa et al. [13]). Outside the optimization community this notion has been already used by Young [27] in algebra, by Nishnianidze [20] in fixed point theory and by Chiriaev and Walster [2] in computer science and interval analysis. Since Chiriaev and Walster introduced the name “set less” for the comparison of sets, which is also implemented

J. Jahn (✉)
Department Mathematik, Universität Erlangen-Nürnberg, Cauerstr. 11, D-91058 Erlangen, Germany
e-mail: johannes.jahn@fau.de

in the FORTRAN compiler f95 of SUN Microsystems [24], we also use this name in the present paper. In [10] even more realistic order relations have been discussed.

Recently, Neukel [19] has shown that there are important socio-economic applications for these order relations. Applications of set optimization to finance can be found in a paper by Hamel and Heyde [6] among others.

This paper aims at the numerical solution of set optimization problems. The presented methods are based on a vectorization approach [8] by the author. Vectorization is a first step to a scalarization in set optimization. Direct scalarization approaches for set optimization problems have been introduced by Hernández and Rodríguez-Marín [7] in 2007 and in 2012 by Gutiérrez et al. [5], Khoshkhabar-amiranloo and Soleimani-damaneh [11] and Xiao et al. [26].

Löhne and Schrage [16] have given the first algorithm for the solution of polyhedral set optimization problems. Based on the vectorization approach a new descent method is now presented for generalized convex optimization problems which can also be used for the treatment of polyhedral sets. It turns out that these nonlinear problems can only be solved with a high numerical effort.

This paper is organized as follows. In Sect. 2, we present the basic theoretical results together with its application to numerical methods for the comparison of two sets with respect to the set less order relation. Using the definition of minimal elements Sect. 3 gives an algorithm for the determination of optimal scenarios. Finally, the descent method is described in Sect. 4 and numerical results are given.

2 Basic results

In this section we investigate set optimization problems in the following setting.

Assumption 2.1 Let S be a nonempty subset of \mathbb{R}^n , let C be a closed convex cone in \mathbb{R}^m with $C \neq \mathbb{R}^m$, and let $F : S \rightrightarrows \mathbb{R}^m$ be a set-valued map with $F(x) \neq \emptyset$, $F(x)$ compact, $F(x) + C$ and $F(x) - C$ convex for all $x \in S$.

Under this assumption we investigate the finite dimensional set optimization problem

$$\min_{x \in S} F(x). \quad (1)$$

In this paper minimal solutions of problem (1) are defined using the set less order relation.

Definition 2.1 Let Assumption 2.1 be satisfied.

- (a) Let nonempty sets $A, B \in \mathbb{R}^m$ be given and let \leq be a partial ordering induced by the convex cone C . Then the *set less* order relation \preceq_S is defined by

$$A \preceq_S B \quad :\iff \quad (\forall a \in A \exists b \in B : a \leq b) \text{ and } (\forall b \in B \exists a \in A : a \leq b). \quad (2)$$

- (b) $\bar{x} \in S$ is called a *minimal solution* of the set optimization problem (1) iff $F(\bar{x})$ is a minimal element of the system of sets $F(x)$ (with arbitrary $x \in S$), i.e.

$$F(x) \preceq_S F(\bar{x}), \quad x \in S \implies F(\bar{x}) \preceq_S F(x).$$

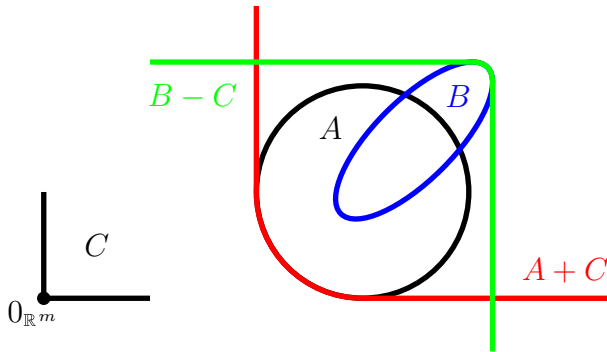


Fig. 1 Illustration of the inequality $A \preceq_s B$, i.e. $B \subset A + C$ and $A \subset B - C$

In the equivalence (2) one requires that two conditions are satisfied in order to have some kind of symmetry in the statement. It is well-known (e.g., see [10, Proposition 3.1,(a)]) that the equivalence (2) can be replaced by

$$A \preceq_s B \iff B \subset A + C \text{ and } A \subset B - C.$$

Figure 1 illustrates these two inclusions used in the characterization of the set less order relation \preceq_s .

Throughout this paper let C^* denote the dual cone of C . Lower indices are used for the components of a vector in \mathbb{R}^n (or \mathbb{R}^m) and different vectors in \mathbb{R}^n (or \mathbb{R}^m) are distinguished by upper indices. All computations in this paper have been done with MATLAB.

For the comparison of two sets using the set less order relation we cite Theorem 2.1 in [8] in a specific form in finite dimensions.

Theorem 2.1 [8, Thm. 2.1] *Under Assumption 2.1 we have for all $x^1, x^2 \in S$*

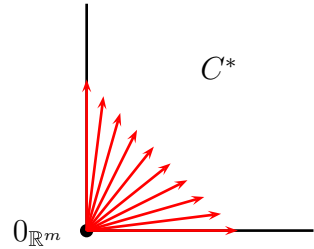
$$F(x^1) \preceq_s F(x^2) \iff \max \left(\sup_{\ell \in C^* \setminus \{0_{\mathbb{R}^m}\}} \left(\min_{y \in F(x^1)} \ell^T y - \min_{y \in F(x^2)} \ell^T y \right), \sup_{\ell \in C^* \setminus \{0_{\mathbb{R}^m}\}} \left(\max_{y \in F(x^1)} \ell^T y - \max_{y \in F(x^2)} \ell^T y \right) \right) \leq 0. \quad (3)$$

We use this equivalence in order to decide on a computer whether $F(x^1) \preceq_s F(x^2)$ or not. Theorem 2.1 works with infinitely many vectors ℓ in the nontrivial dual cone and this result can be reduced to finitely many vectors ℓ in special cases.

2.1 The nonlinear case

It is remarked in [8, Remark 2.2] that the complexity of the set $C^* \setminus \{0_{\mathbb{R}^m}\}$ is reduced, if it is restricted to the subset $C^* \cap \{y \in \mathbb{R}^m \mid \|y\| = 1\}$ belonging to the unit sphere where $\|\cdot\|$ is an arbitrary norm in \mathbb{R}^m . Although this set also consists of infinitely

Fig. 2 Discretization of C^*



many elements it can be discretized in small dimensions. Figure 2 illustrates such a discretization for $m = 2$ and $C^* = \mathbb{R}_+^2$.

If we replace the dual cone C^* by such a subset of finitely many vectors $\ell^1, \dots, \ell^s \in \mathbb{R}^m$ for a given $s \in \mathbb{N}$, then the equivalence (3) in Theorem 2.1 can be replaced by the approximate equivalence

$$F(x^1) \preceq_s F(x^2) \iff \max \left(\max_{\ell \in \{\ell^1, \dots, \ell^s\}} \left(\min_{y \in F(x^1)} \ell^T y - \min_{y \in F(x^2)} \ell^T y \right), \max_{\ell \in \{\ell^1, \dots, \ell^s\}} \left(\max_{y \in F(x^1)} \ell^T y - \max_{y \in F(x^2)} \ell^T y \right) \right) \leq 0. \quad (4)$$

For a given discretization $\{\ell^1, \dots, \ell^s\}$ of the intersection of the unit sphere and the dual cone the inequality in (4) can be checked on a computer. One has to solve $4s$ nonlinear optimization problems. Since these optimization problems have the same constraint sets $F(x^1)$ and $F(x^2)$ and only the linear objective functions vary, it makes certainly sense to solve these $4s$ subproblems in parallel. So, the decision, whether $F(x^1) \preceq_s F(x^2)$ or not, can be made by highly parallelized computations.

In the following remark we investigate a possible error bound caused by the discretization of the set $C^* \cap \{y \in \mathbb{R}^m \mid \|y\| = 1\}$.

Remark 2.1 For convenience we set $T := C^* \cap \{y \in \mathbb{R}^m \mid \|y\| = 1\}$, where the norm $\|\cdot\|$ now denotes the Euclidean norm, and for an arbitrary $x \in S$ we define the minimal value function $\varphi_{min}^x : T \rightarrow \mathbb{R}$ with

$$\varphi_{min}^x(\ell) := \min_{y \in F(x)} \ell^T y \text{ for all } \ell \in T.$$

For simplicity we investigate only one inner term in the max term in (4). For arbitrarily given $x^1, x^2 \in S$ we define the function $\varphi^{x^1, x^2} : T \rightarrow \mathbb{R}$ with

$$\varphi^{x^1, x^2}(\ell) := \varphi_{min}^{x^1}(\ell) - \varphi_{min}^{x^2}(\ell) \text{ for all } \ell \in T.$$

By Assumption 2.1 the constraint sets $F(x^1)$ and $F(x^2)$ are compact and because the objective functions are linear, the minimal value functions $\varphi_{min}^{x^1}$ and $\varphi_{min}^{x^2}$ are continuous (e.g., see [9, Lemma 2.1]) and, therefore, the function φ^{x^1, x^2} is continuous as well. Since the set T is compact, there is some $\bar{\ell} \in T$ with

$$\sup_{\ell \in T} \left(\min_{y \in F(x^1)} \ell^T y - \min_{y \in F(x^2)} \ell^T y \right) = \sup_{\ell \in T} \varphi^{x^1, x^2}(\ell) = \max_{\ell \in T} \varphi^{x^1, x^2}(\ell) = \varphi^{x^1, x^2}(\bar{\ell}). \tag{5}$$

If we consider a discretization $\{\ell^1, \dots, \ell^s\}$ (with $s \in \mathbb{N}$) of the set T , we get for some ℓ^j with $j \in \{1, \dots, s\}$

$$\max_{\ell \in \{\ell^1, \dots, \ell^s\}} \left(\min_{y \in F(x^1)} \ell^T y - \min_{y \in F(x^2)} \ell^T y \right) = \max_{\ell \in \{\ell^1, \dots, \ell^s\}} \varphi^{x^1, x^2}(\ell) = \varphi^{x^1, x^2}(\ell^j). \tag{6}$$

Let ℓ^k with $k \in \{1, \dots, s\}$ be a point of the discrete set $\{\ell^1, \dots, \ell^s\}$ with smallest distance to the point $\bar{\ell}$ and let for every $\ell \in T$ the smallest distance to a point ℓ^1, \dots, ℓ^s be bounded by a number $d > 0$ describing the fineness of discretization. If the sets $F(x^1)$ and $F(x^2)$ are given by inequality and equality constraints, it can be shown under strong assumptions (e.g., appropriate differentiability assumptions, the strong second-order sufficient condition and the linear independence constraint qualification) that the minimal value functions $\varphi_{min}^{x^1}$ and $\varphi_{min}^{x^2}$ are differentiable (for a similar result see [9, Lemma 4.1]). If the norm of the gradients $\nabla \varphi_{min}^{x^1}$ and $\nabla \varphi_{min}^{x^2}$ are bounded on the set T by some $u > 0$, then by the mean value theorem there is some $\vartheta \in (0, 1)$ so that with the equalities (5) and (6)

$$\begin{aligned} & \left| \sup_{\ell \in T} \left(\min_{y \in F(x^1)} \ell^T y - \min_{y \in F(x^2)} \ell^T y \right) - \max_{\ell \in \{\ell^1, \dots, \ell^s\}} \left(\min_{y \in F(x^1)} \ell^T y - \min_{y \in F(x^2)} \ell^T y \right) \right| \\ &= |\varphi^{x^1, x^2}(\bar{\ell}) - \varphi^{x^1, x^2}(\ell^j)| \\ &\leq |\varphi^{x^1, x^2}(\bar{\ell}) - \varphi^{x^1, x^2}(\ell^k)| \\ &= |\nabla \varphi^{x^1, x^2}(\ell^k + \vartheta(\bar{\ell} - \ell^k))^T (\bar{\ell} - \ell^k)| \\ &= |(\nabla \varphi_{min}^{x^1}(\ell^k + \vartheta(\bar{\ell} - \ell^k)) - \nabla \varphi_{min}^{x^2}(\ell^k + \vartheta(\bar{\ell} - \ell^k)))^T (\bar{\ell} - \ell^k)| \\ &\leq (\|\nabla \varphi_{min}^{x^1}(\ell^k + \vartheta(\bar{\ell} - \ell^k))\| + \|\nabla \varphi_{min}^{x^2}(\ell^k + \vartheta(\bar{\ell} - \ell^k))\|) \|\bar{\ell} - \ell^k\| \\ &\leq 2u \cdot d. \end{aligned}$$

Recall that d is a measure of the fineness of discretization of the set T . So, the obtained error bound decreases, if the fineness of discretization is increased.

2.2 The polyhedral case

In this subsection we restrict ourselves to polyhedral sets $F(x)$ (with $x \in S$). In this case the result of Theorem 2.1 can also be reduced to finitely many vectors $\ell \in C^* \setminus \{0_{\mathbb{R}^m}\}$.

Theorem 2.2 *Let Assumption 2.1 be satisfied. In addition, let the cone C be polyhedral and let the set-valued map F be given with*

$$F(x) := \{y \in \mathbb{R}^m \mid A(x) \cdot y \leq b(x)\} \text{ for all } x \in S$$

where $A : S \rightarrow \mathbb{R}^{(p, m)}$ and $b : S \rightarrow \mathbb{R}^p$ are given maps. For every $x \in S$ let the set

$$L_1(x) := \{-a^1(x), \dots, -a^p(x)\} \cap (C^* \setminus \{0_{\mathbb{R}^m}\})$$

and

$$L_2 := \{\ell^1, \dots, \ell^s\}$$

be given where $a^1(x), \dots, a^p(x)$ denote the rows of the matrix $A(x)$ (with $x \in S$) and ℓ^1, \dots, ℓ^s are the normed extremal points of the dual cone C^* . Then for all $x^1, x^2 \in S$ it holds

$$F(x^1) \preceq_S F(x^2) \iff \forall \ell \in L_1(x^1) : b_t(x^1) \geq \max_{y \in F(x^2)} -\ell^T y \tag{7}$$

(for $t \in \{1, \dots, p\}$ with $\ell = -a^t(x^1)$)

$$\forall \ell \in L_2 : \min_{y \in F(x^1)} \ell^T y \leq \min_{y \in F(x^2)} \ell^T y$$

and

$$\forall \ell \in L_1(x^2) \cup L_2 : \max_{y \in F(x^1)} \ell^T y \leq \max_{y \in F(x^2)} \ell^T y$$

$$\iff \max \left(\max_{\ell \in L_1(x^1)} (-b_t(x^1) + \max_{y \in F(x^2)} -\ell^T y), \right.$$

$$\left. \max_{\ell \in L_2} \left(\min_{y \in F(x^1)} \ell^T y - \min_{y \in F(x^2)} \ell^T y \right), \right.$$

$$\left. \max_{\ell \in L_1(x^2) \cup L_2} \left(\max_{y \in F(x^1)} \ell^T y - \max_{y \in F(x^2)} \ell^T y \right) \right) \leq 0 \tag{8}$$

(for $t \in \{1, \dots, p\}$ with $\ell = -a_t(x^1)$).

Proof The proof of the second equivalence is obvious. The proof of the first equivalence consists of two parts: an introductory part and the proof of this equivalence. Let $x^1, x^2 \in S$ be arbitrarily chosen.

(a) For an arbitrary $\ell \in L_1(x^1)$ we have with the definition of the set $F(x^1)$

$$b_t(x^1) = \max_{y \in F(x^1)} a^t(x^1)^T y = - \min_{y \in F(x^1)} -a^t(x^1)^T y = - \min_{y \in F(x^1)} \ell^T y$$

(for $t \in \{1, \dots, p\}$ with $\ell = -a^t(x^1)$) and, therefore, the inequality (7) is equivalent to

$$\min_{y \in F(x^1)} \ell^T y \leq \min_{y \in F(x^2)} \ell^T y.$$

Thus, the right hand side of the equivalence in the assertion can be written as

$$\forall \ell \in L_1(x^1) \cup L_2 : \min_{y \in F(x^1)} \ell^T y \leq \min_{y \in F(x^2)} \ell^T y$$

and

$$\forall \ell \in L_1(x^2) \cup L_2 : \max_{y \in F(x^1)} \ell^T y \leq \max_{y \in F(x^2)} \ell^T y.$$

(b) For the actual proof of the first equivalence we notice that the implication “ \implies ” is obvious with Theorem 2.1 and part (a) of this proof. Finally, we prove the opposite direction “ \impliedby ”. It is well-known that the polyhedron $F(x^1) + C$ can be represented by the intersection of certain affine halfspaces generated by facets of this polyhedron, i.e. $F(x^1) + C = \bigcap_{\ell \in L_1(x^1) \cup L_2} H_1(\ell)$ for appropriate affine halfspaces $H_1(\ell)$ ($\ell \in L_1(x^1) \cup L_2$). Because of the assumption and the equivalent transformation in part (a) of this proof we obtain

$$F(x^2) \subset H_1(\ell) \text{ for all } \ell \in L_1(x^1) \cup L_2$$

which implies

$$F(x^2) \subset \bigcap_{\ell \in L_1(x^1) \cup L_2} H_1(\ell) = F(x^1) + C. \tag{9}$$

With the same arguments one can consider the affine halfspaces generated by facets of the polyhedron $F(x^2) - C$, one can apply part (a) of this proof and then one gets

$$F(x^1) \subset F(x^2) - C. \tag{10}$$

From the inclusions (9) and (10) it follows with Definition 2.1,(a) that $F(x^1) \preceq_s F(x^2)$ which has to be shown. \square

Remark 2.2 The result of Theorem 2.2 has an important consequence from a numerical point of view. The decision whether $F(x^1) \preceq_s F(x^2)$ can be made by solving a finite number of linear optimization problems. The number of linear problems depends on the number of appropriate facets of the polyhedrons $F(x^1) + C$ and $F(x^2) - C$. So, for problems with a less complex polyhedral structure this number is small.

The result of Theorem 2.2 can be used in an algorithm in order to evaluate whether $F(x^1) \preceq_s F(x^2)$ or $F(x^1) \not\preceq_s F(x^2)$.

Algorithm 2.1 (Test of $F(x^1) \preceq_s F(x^2)$ for $x^1, x^2 \in S$)

```

Input:  $x^1, x^2 \in S, F : S \rightrightarrows \mathbb{R}^m$ 
 $L_1(x^1) := \{-a^1(x^1), \dots, -a^p(x^1)\} \cap (C^* \setminus \{0_{\mathbb{R}^m}\})$ 
 $L_1(x^2) := \{-a^1(x^2), \dots, -a^p(x^2)\} \cap (C^* \setminus \{0_{\mathbb{R}^m}\})$ 
 $L_2 := \{\ell^1, \dots, \ell^s\}$ 
for  $t = 1 : 1 : \#(L_1(x^1))$  do
     $\ell^t \in L_1(x^1)$ 
     $\alpha_1 := \max_{A(x^2)y \leq b(x^2)} -\ell^{tT}y$ 
    if  $b_t(x^1) < \alpha_1$  then  $F(x^1) \preceq_s F(x^2)$  is not true, stop
    end if
end for
for  $t = 1 : 1 : s$  do
     $\ell^t \in L_2$ 
     $\alpha_2 := \min_{A(x^1)y \leq b(x^1)} \ell^{tT}y$ 

```

```

 $\alpha_3 := \min_{A(x^2)y \leq b(x^2)} \ell^T y$ 
if  $\alpha_2 > \alpha_3$  then  $F(x^1) \preceq_s F(x^2)$  is not true, stop
end if
end for
for  $t = 1 : 1 : \#(L_1(x^2)) + s$  do
 $\ell^t \in L_1(x^2) \cup L_2$ 
 $\alpha_4 := \max_{A(x^1)y \leq b(x^1)} \ell^t y$ 
 $\alpha_5 := \max_{A(x^2)y \leq b(x^2)} \ell^t y$ 
if  $\alpha_4 > \alpha_5$  then  $F(x^1) \preceq_s F(x^2)$  is not true, stop
end if
end for
 $F(x^1) \preceq_s F(x^2)$  is true, stop
Output:  $F(x^1) \preceq_s F(x^2)$  or  $F(x^1) \not\preceq_s F(x^2)$ 

```

The following corollary is a direct consequence of Theorem 2.2.

Corollary 2.1 *Under the assumptions of Theorem 2.2 Algorithm 2.1 is well-defined and makes the decision whether $F(x^1) \preceq_s F(x^2)$ is true or not (for arbitrary $x^1, x^2 \in S$).*

Example 2.1 We investigate the set-valued map $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ with

$$F(x) := \{y \in \mathbb{R}^2 \mid A(x) \cdot y \leq b(x)\} \forall x \in \mathbb{R}^2$$

where for arbitrary $x = (x_1, x_2) \in \mathbb{R}^2$

$$A(x) := \begin{pmatrix} -\frac{1}{1+|x_2-x_1|} & -\frac{1}{3+|x_2-x_1|} \\ -\frac{1}{1+|x_1|} & -x_2 \\ -\frac{5}{4} & 1 \\ \frac{6}{5} & -1 \\ \frac{5-|x_2-x_1|}{7} & 1 \end{pmatrix} \quad \text{and} \quad b(x) := \begin{pmatrix} -1 \\ -1 \\ 2+x_2 \\ \frac{41}{5} - \frac{16}{25}(2x_1-x_2) \\ 9+2x_1 - \frac{2}{35}(x_1-3x_2) \end{pmatrix}.$$

Figure 3 illustrates the sets $F(1, 2)$ and $F(3, 1)$. For the cone $C := \mathbb{R}_+^2$ we obtain with Algorithm 2.1 the result $F(1, 2) \preceq_s F(3, 1)$.

3 Determination of optimal scenarios

In concrete applications, like socio-economic applications discussed by Neukel in [19], the constraint set S only consists of some few points $\{x^1, \dots, x^k\} \subset \mathbb{R}^n$ (with $k \in \mathbb{N}$). Every set $F(x^i)$ (with $i \in \{1, \dots, k\}$) describes a certain scenario and it is the aim to find optimal scenarios. This means mathematically that one has to determine

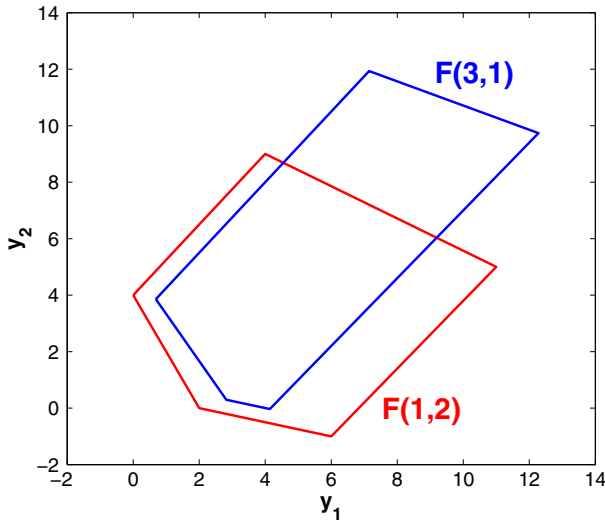


Fig. 3 Illustration of the sets $F(1, 2)$ and $F(3, 1)$ in Example 2.1

minimal elements of the system of sets $F(x^i)$ with $i \in \{1, \dots, k\}$, if they exist. Based on the definition of minimal solutions the following algorithm determines optimal scenarios.

Algorithm 3.1 (Determination of minimal solutions)

Input: $S := \{x^1, \dots, x^k\}$, $F : S \rightrightarrows \mathbb{R}^m$

```

for  $i = 1 : 1 : k$  do
    for  $j = 1 : 1 : k$  with  $j \neq i$  do
        if  $F(x^j) \preceq_s F(x^i)$  then
            if  $F(x^i) \not\preceq_s F(x^j)$  then  $x^i$  is not a minimal solution
                and continue with “for 1”
            end if
        end if
    end for 2
     $x^i$  minimal solution
end for 1
    
```

Output: minimal solutions, if they exist

Proposition 3.1 *Let Assumption 2.1 be satisfied. Algorithm 3.1 is well-defined and determines all minimal solutions, if there exists at least one minimal solution.*

Proof The proof is obvious because the algorithm works according to the definition of minimal solutions. □

Recently, Löhne [15] has proved by induction that every finite set S has at least one minimal solution so that the corresponding assumption in Proposition 3.1 can be dropped.

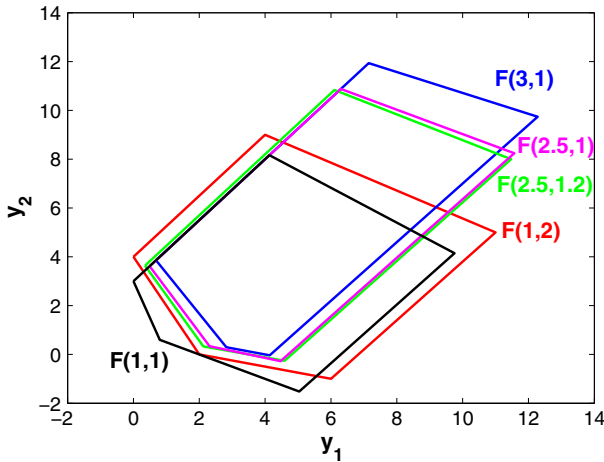


Fig. 4 Illustration of the sets $F(1, 1)$, $F(1, 2)$, $F(2.5, 1)$, $F(2.5, 1.2)$ and $F(3, 1)$ in Example 3.1

Remark 3.1 In the nonlinear case the inequality in the approximate equivalence (4) can be used for the comparison of sets in Algorithm 3.1. In the polyhedral case one works with Algorithm 2.1 for the comparisons in the “if” parts of Algorithm 3.1.

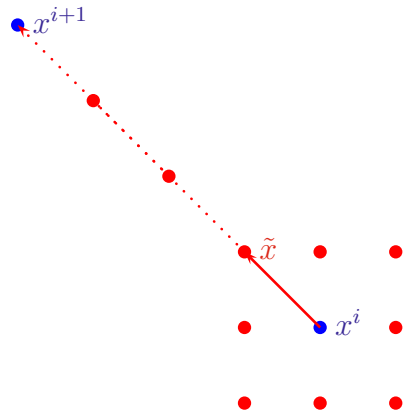
As a simple example we consider a problem of finding an optimal scenario among given scenarios.

Example 3.1 For the set-valued map $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ defined in Example 2.1 we consider the sets $F(1, 1)$, $F(1, 2)$, $F(2.5, 1)$, $F(2.5, 1.2)$ and $F(3, 1)$. These sets are illustrated in Fig. 4. It is obvious from this figure that for the natural order cone there is only one minimal scenario among these 5 scenarios. In other words, for $C = \mathbb{R}_+^2$ and $S := \{(1, 1), (1, 2), (2.5, 1), (2.5, 1.2), (3, 1)\}$ the point $(1, 1)$ is a minimal solution of problem (1). This result is obtained with Algorithm 3.1.

4 A descent method

Until now we have investigated set optimization problems with a constraint set consisting of a finite number of points. In this section we now consider unconstrained set optimization problems, i.e. in Assumption 2.1 with $S = \mathbb{R}^n$. We present a derivative-free descent method which works with a cluster of points. For instance, the well-known Nelder–Mead method (see [14, 17, 18]) is a classical method in nonlinear optimization which works with a cluster of points. In general, the Nelder–Mead method does not work in set optimization because $n + 1$ points in a cluster in \mathbb{R}^n are too less for comparisons of sets and it does not make sense to drop undesirable points since many sets are not comparable with respect to the set less order relation. Using a cluster of points we now construct a descent direction together with an appropriate step length.

Fig. 5 Illustration of an iteration step of the descent method



Remark 4.1 Let $x^i \in \mathbb{R}^2$ be an arbitrary iteration point. As indicated in Fig. 5 eight equally distributed points with equal small ℓ_∞ distance to x^i are considered and for each point x of this cluster it is checked using the approximate equivalence (4) or the equivalence (8) whether $F(x) \preceq_s F(x^i)$ or not. Among all cluster points satisfying this inequality one chooses the point for which the exact or approximated max term in (4) or (8) has the smallest value. For such a point \tilde{x} the difference $h := \tilde{x} - x^i$ gives the descent direction of this iteration. For the determination of the step length one considers the point $x^i + 2h$. If $F(x^i + 2h) \not\preceq_s F(\tilde{x})$, then one stops and $x^{i+1} := \tilde{x}$ is the new iteration point. Otherwise one considers the point $x^i + 3h$ and so on. This technique is implemented in Algorithm 4.1.

We now present an algorithm for the improvement of feasible points.

Algorithm 4.1 (descent method)

Input: $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, starting point $x^0 \in \mathbb{R}^n$, maximal number i_{max} of iterations, maximal number j_{max} of iterations for the determination of a step length, number $k \in \mathbb{N}$ of points in the cluster, radius $\varepsilon > 0$ of a small ball, discretization multiplier $d \in \mathbb{N}$ with $d > 1$.

% initialization

choose k points $\tilde{x}^1, \dots, \tilde{x}^k$ on the sphere around $0_{\mathbb{R}^n}$ with radius ε

choose $d \cdot k$ points $\hat{x}^1, \dots, \hat{x}^{d \cdot k}$ on the sphere around $0_{\mathbb{R}^n}$ with radius $\frac{\varepsilon}{d}$

$i := 0$

% iteration loop

while $i \leq i_{max}$ **do**

check $F(x^i + \tilde{x}^j) \preceq_s F(x^i)$ for all $j \in \{1, \dots, k\}$ by calculating the max term in (3) (using the approximation (4) in the nonlinear case or the term (8) in the polyhedral case) and determine the point $\tilde{x} \in \{\tilde{x}^1, \dots, \tilde{x}^k\}$ with the smallest max term max_{min}

if $max_{min} \leq 0$ **then**

$h^i := \tilde{x} - x^i$ *% descent direction*

$x^{i+1} := x^i + h^i$ *% new iteration point with step length 1*

$j := 1$

```

while  $F(x^i + (j + 1)h^i) \preceq_s F(x^i + jh^i)$  and  $j \leq j_{max}$  do
     $j := j + 1$ 
     $\lambda^i := j$  % step length
     $x^{i+1} := x^i + \lambda^i h^i$  % new iteration point
end while
else
    check  $F(x^i + \hat{x}^j) \preceq_s F(x^i)$  for all  $j \in \{1, \dots, d \cdot k\}$  by calculating the max
    term in (3) (using the approximation (4) in the nonlinear case or the term
    (8) in the polyhedral case) and determine the point  $\hat{x} \in \{\hat{x}^1, \dots, \hat{x}^{d \cdot k}\}$  with
    the smallest max term  $max_{min}$ 
    if  $max_{min} \leq 0$  then
         $x^{i+1} := x^i + \hat{x}$ 
    else
        stop
    end if
end if
     $i := i + 1$ 
end while
Output:  $x^i$ 

```

It should be noted that in the “else” part of the “if” statement of the preceding algorithm, more than only one refinement of the grid should be implemented in order to have a better chance to find a new iteration point.

Theorem 4.1 *Let Assumption 2.1 with $S = \mathbb{R}^n$ be satisfied.*

- (a) *Algorithm 4.1 is well-defined for arbitrary input parameters.*
- (b) *For an arbitrary starting point $x^0 \in \mathbb{R}^n$ let $x^1, x^2, \dots \in \mathbb{R}^n$ denote the iteration points generated by Algorithm 4.1.*
 - (i) *In the polyhedral case we have*

$$F(x^{i+1}) \preceq_s F(x^i) \text{ for all } i = 0, 1, 2, \dots \tag{11}$$

This inequality is also satisfied in the nonlinear case, if one works with all $\ell \in C^ \setminus \{0_{\mathbb{R}^m}\}$.*

- (ii) *If $x' \in \mathbb{R}^n$ denotes the output vector, then we obtain in the polyhedral case*

$$F(x') \preceq_s F(x^0).$$

This inequality is also satisfied in the nonlinear case, if one works with all $\ell \in C^ \setminus \{0_{\mathbb{R}^m}\}$.*

- (c) *For arbitrary input parameters where i_{max} and j_{max} are sufficiently large, let $S(0_{\mathbb{R}^m}, \frac{\epsilon}{d})$ denote the sphere around $0_{\mathbb{R}^m}$ with radius $\frac{\epsilon}{d}$. If one works with the whole sphere $S(0_{\mathbb{R}^m}, \frac{\epsilon}{d})$ in Algorithm 4.1 (and not only with a cluster of points), then in the polyhedral case the output vector $x' \in \mathbb{R}^n$ is minimal with respect to the set $S(x', \frac{\epsilon}{d}) \cup \{x'\}$. This assertion is also true in the nonlinear case, if one works with all $\ell \in C^* \setminus \{0_{\mathbb{R}^m}\}$.*

Proof (a) This part of the proof is obvious.

- (b),(i) For an arbitrary iteration point $x^i \in \mathbb{R}^n$ Algorithm 4.1 determines a new iteration point x^{i+1} , if the corresponding max term is nonpositive. In the polyhedral case we then get $F(x^{i+1}) \preceq_s F(x^i)$ by Theorem 2.2. In the nonlinear case we get the assertion using Theorem 2.1.
- (b),(ii) This assertion immediately follows from part (i) because the set less order relation \preceq_s is transitive.
- (c) If Algorithm 4.1 stops at an iteration point $x' \in \mathbb{R}^n$, then the method could not find any $x \in S(x', \frac{\varepsilon}{d})$ with the property $F(x) \preceq_s F(x')$. So, by Definition 2.1,(b) x' is a minimal solution with respect to the set $S(x', \frac{\varepsilon}{d}) \cup \{x'\}$.

□

Remark 4.2

- (a) Because of the inequality (11) we use the concept *descent method* for Algorithm 4.1.
- (b) The assertion (c) in Theorem 4.1 says that the output vector x' is some kind of minimal solution. Since $\varepsilon > 0$ is small and $d > 1$, this means some kind of local minimality. At the best, the vector x' is a local minimal solution.
- (c) In practice, Algorithm 4.1 does not produce locally minimal solutions because various discretizations are included in this method. In the nonlinear case only finitely many vectors are chosen in $C^* \setminus \{0_{\mathbb{R}^m}\}$ and in both cases Algorithm 4.1 works only with finitely many vectors in the spheres $S(0_{\mathbb{R}^m}, \varepsilon)$ and $S(0_{\mathbb{R}^m}, \frac{\varepsilon}{d})$. If these discretizations are not fine enough, it may be that descent directions cannot be found. The radius $\varepsilon > 0$ should be small enough in order to obtain local descent directions.

Algorithm 4.1 fits into the class of pattern search methods with one essential difference that we do not minimize an objective function but we work with a function $\varphi : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ with

$$\varphi(\bar{x}, \tilde{x}) = \max \left(\sup_{\ell \in C^* \setminus \{0_{\mathbb{R}^m}\}} \left(\min_{y \in F(\bar{x})} \ell^T y - \min_{y \in F(\tilde{x})} \ell^T y \right), \right. \\ \left. \sup_{\ell \in C^* \setminus \{0_{\mathbb{R}^m}\}} \left(\max_{y \in F(\bar{x})} \ell^T y - \max_{y \in F(\tilde{x})} \ell^T y \right) \right) \text{ for all } \bar{x}, \tilde{x} \in \mathbb{R}^n$$

which has a special meaning in this setting. In fact, we want to obtain negative function values which should finally tend to zero. Torczon [25] has developed a general convergence theory for pattern search algorithms which can be adapted to the special problem in this paper (see also [3]).

In order to obtain such a convergence result we have to extend Algorithm 4.1 with two additional specifications (in the following we set $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$):

- (A) Assume that the pattern contains at least one direction of descent whenever a set $F(x^i)$ ($i \in \mathbb{N}_0$) can be improved.
- (B) Let some $\beta \in (0, 1)$ and an arbitrary null sequence $(\varepsilon^i)_{i \in \mathbb{N}_0}$ with $\varepsilon^i < 0$ for all $i \in \mathbb{N}_0$ be given. While $\varphi(x^i + \lambda h^i, x^i) \leq \varepsilon^i < 0$ set $\lambda := \beta^q \lambda^0$ for $q := 0, 1, 2, \dots$ (λ^0 and h^i denote the initial step length and the descent direction at an iteration point x^i , respectively).

If an arbitrary point x^i ($i \in \mathbb{N}_0$) is a non-final iteration point, then there exists a (strict) descent direction and an $\bar{x} \in \mathbb{R}^n$ with $\varphi(\bar{x}, x^i) < 0$. Since the function φ is continuous, there exists a ball $B(\bar{x}, \delta)$ around \bar{x} with radius $\delta > 0$ so that

$$\varphi(x, x^i) < 0 \text{ for all } x \in B(\bar{x}, \delta).$$

So, it is obvious that a refinement of the grid leads to a descent direction so that the specification (A) is satisfied.

Specification (B) defines a step length control where various forms are possible. It should be ensured that the step lengths tend to zero. Since the function φ is continuous, it is always possible to give some finite $q \in \mathbb{N}$ so that for $\lambda := \beta^q \lambda^0$ the inequality $0 \geq \varphi(x^i + \lambda h^i, x^i) > \varepsilon^i$ holds (if $\varphi(x^i + \lambda h^i, x^i) \leq 0$ for all $\lambda \in [0, 1]$), i.e. the while loop in specification (B) can be carried out in finitely many steps.

With these two specifications we now formulate a convergence result.

Theorem 4.2 *Let Assumption 2.1 be satisfied, let Algorithm 4.1 with the additional specifications (A) and (B) generate an iteration sequence $(x^i)_{i \in \mathbb{N}_0}$ and let the level set*

$$L_{x^0} := \left\{ x \in \mathbb{R}^n \mid F(x) \preceq_s F(x^0) \right\}$$

be compact. Then it follows

$$\limsup_{i \rightarrow \infty} \varphi(x^{i+1}, x^i) = 0.$$

Proof Since Algorithm 4.1 is a descent method, we have $x^i \in L_{x^0}$ for all $i \in \mathbb{N}_0$. Because of the compactness of the level set L_{x^0} and the continuity of the function φ (compare Remark 2.1), the function values $\varphi(x^{i+1}, x^i)$ with $i \in \mathbb{N}_0$ are bounded. Consequently, the limit superior exists.

Assume that $\limsup \varphi(x^{i+1}, x^i) \neq 0$. Then there exists a subsequence $(x^{i_r})_{r \in \mathbb{N}}$ with $\lim_{r \rightarrow \infty} \varphi(x^{i_r+1}, x^{i_r}) := \alpha \neq 0$. With the specification (A) we have $F(x^{i+1}) \preceq_s F(x^i)$ for all $i \in \mathbb{N}_0$ and then we get by Theorem 2.1

$$\varphi(x^{i+1}, x^i) \leq 0 \text{ for all } i \in \mathbb{N}_0.$$

Consequently, we have $\alpha < 0$. Then there is some $N_1 \in \mathbb{N}$ with the property

$$\varphi(x^{i_r+1}, x^{i_r}) \leq \frac{\alpha}{2} < 0 \text{ for all } r \geq N_1.$$

Since $(\varepsilon^i)_{i \in \mathbb{N}_0}$ is a null sequence, there is some $N_2 \in \mathbb{N}$ with

$$\frac{\alpha}{2} \leq \varepsilon^{i_r} < 0 \text{ for all } r \geq N_2.$$

Then we obtain

$$\varphi(x^{i_r+1}, x^{i_r}) \leq \frac{\alpha}{2} \leq \varepsilon^{i_r} \text{ for all } r \geq N := \max\{N_1, N_2\},$$

but this contradicts the specification (B). □

In the following Algorithm 4.1 is applied to various examples in \mathbb{R}^2 .

Example 4.1 We investigate the set-valued map $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ with

$$F(x) := \{y \in \mathbb{R}^2 \mid A(x) \cdot y \leq b(x)\} \forall x \in \mathbb{R}^2$$

where for arbitrary $x = (x_1, x_2) \in \mathbb{R}^2$

$$A(x) := \begin{pmatrix} -\frac{1}{1+|x_2|} & -\frac{1}{3+|x_1|} \\ -\frac{1}{1+|x_1|} & -x_2 \\ -\frac{5}{4} & 1 \\ \frac{6}{5} & -1 \\ \frac{5-|x_2|}{7} & 1 \end{pmatrix} \text{ and } b(x) := \begin{pmatrix} -1 \\ -1 \\ 2+x_2 \\ \frac{41}{5} - \frac{16}{25}(2x_1-x_2) \\ 9+2x_1 - \frac{2}{35}(x_1-3x_2) \end{pmatrix}.$$

The matrix function A is slightly different from the one in Example 2.1. Figure 6 illustrates the iteration process for $C := \mathbb{R}_+^2$. The chosen starting point is $x^0 := (3, 2.5)$, the radius is $\varepsilon := 0.01$ and $k := 8$ cluster points are selected as indicated in Fig. 5. Algorithm 4.1 then determines the iteration points $x^1 := (1.34, 0.84)$, $x^2 := (0.71, 0.84)$ and $x^3 := (0.18, 0.31)$. Using a discretization multiplier $d := 2$ the point $(0.175, 0.3075)$ cannot be improved. For this example totally 4,566 linear optimization problems have to be solved. Figure 7 illustrates the sets $F(x^0)$ and $F(x^3)$.

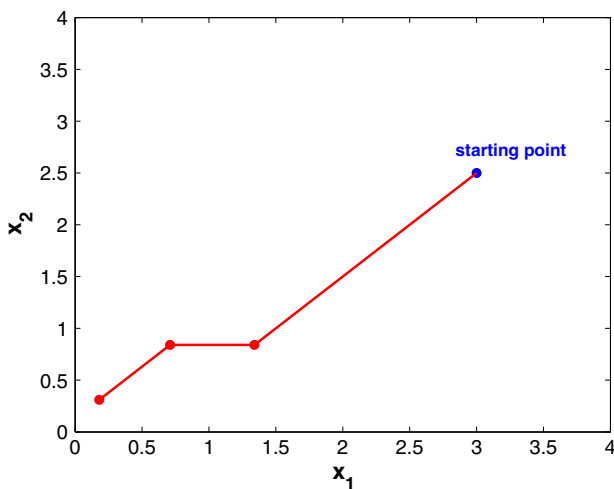


Fig. 6 Illustration of the iteration points in Example 4.1

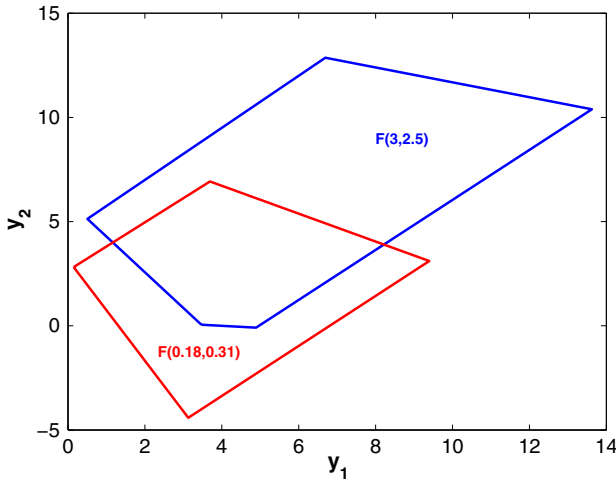


Fig. 7 Illustration of the sets $F(3, 2.5)$ and $F(0.18, 0.31)$ in Example 4.1

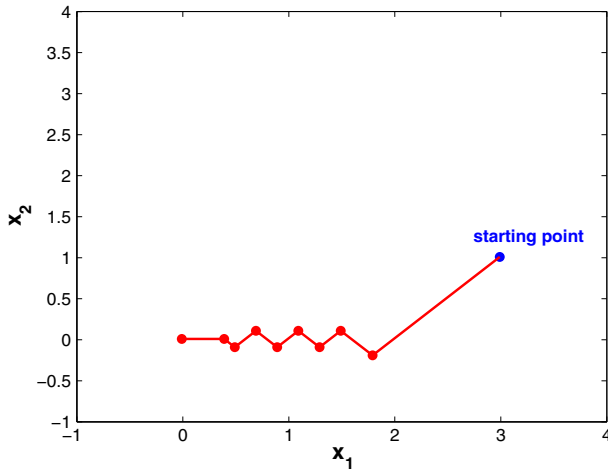


Fig. 8 Illustration of the iteration points in Example 4.2

Example 4.2 We investigate the set-valued map $F : \mathbb{R}^2 \rightrightarrows \mathbb{R}^2$ with

$$F(x_1, x_2) := \left\{ (y_1, y_2) \in \mathbb{R}^2 \mid (y_1 - 2x_1^2)^2 + (y_2 - 2x_2^2)^2 \leq (1 + x_2^2)^2 \right\} \quad \forall (x_1, x_2) \in \mathbb{R}^2.$$

Here we choose the starting point $x^0 := (2.99, 1.01)$, the radius $\varepsilon := 0.1$, the discretization parameter $d := 2$ and the number $s := 10$ of vectors in the dual cone C^* generated by the convex hull of the two points $(0.1111, 0.8889)$ and $(1.1111, -0.1111)$. Figure 8 illustrates the iteration process and Table 1 gives the iteration points obtained by Algorithm 4.1. The iteration point x^9 cannot be improved. For this example totally 3,340 convex optimization problems have to be solved.

Table 1 Iteration points in Example 4.2

$x^1 = (1.79, -0.19)$
$x^2 = (1.49, 0.11)$
$x^3 = (1.29, -0.09)$
$x^4 = (1.09, 0.11)$
$x^5 = (0.89, -0.09)$
$x^6 = (0.69, 0.11)$
$x^7 = (0.49, -0.09)$
$x^8 = (0.39, 0.01)$
$x^9 = (-0.01, 0.01)$

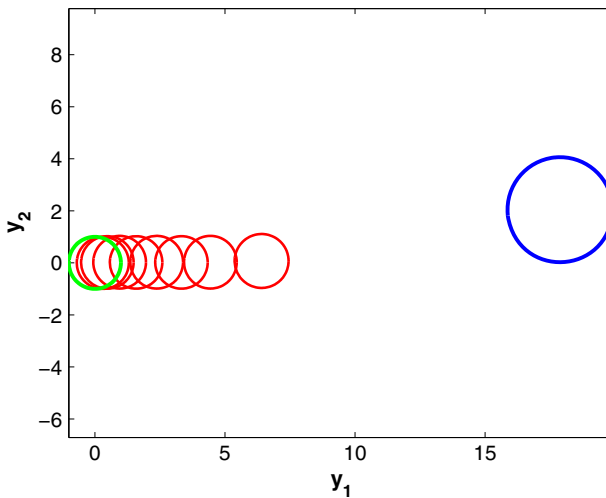


Fig. 9 Illustration of the sets $F(x^0), \dots, F(x^9)$ in Example 4.2 ($F(x^0)$ is the set on the right and $F(x^9)$ is the left circle)

The solution of these convex optimization problems has been done using the *fmincon* tool in MATLAB. Figure 9 illustrates the sets $F(x^i)$ for $i = 0, 1, \dots, 9$.

Example 4.3 Again we consider the problem in Example 4.2; but now we work with $C := \mathbb{R}_+^2$ and a higher discretization in the dual cone. Here we choose $s := 100$ vectors in $C^* = \mathbb{R}_+^2$. For the starting point $x^0 := (1, 1.2)$ we get with the parameter $\varepsilon := 0.1$ the iteration point $x^1 := (0.1, 0.3)$ which cannot be improved in this setting. Figure 10 illustrates the sets $F(x^0)$ and $F(x^1)$. With the multiplier $d := 2$ one obtains the slightly changed point $(1.9429 \cdot 10^{-16}, 0.25)$ which cannot be improved. Because of the high discretization totally 21,000 convex optimization problems have to be solved.

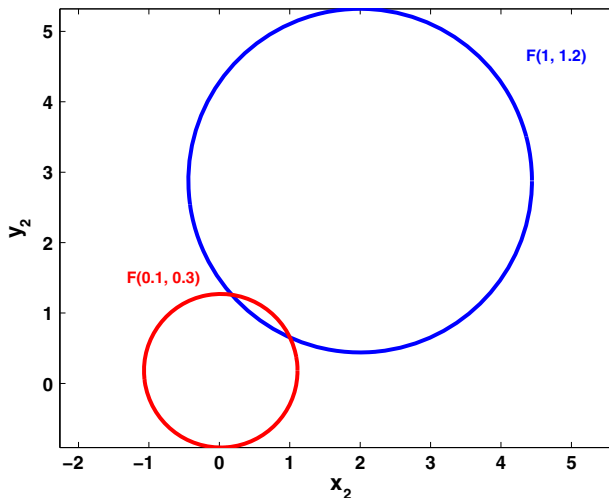


Fig. 10 Illustration of the sets $F(x^0)$ and $F(x^1)$ in Example 4.3

5 Conclusion

The presented descent method in Algorithm 4.1 does not generally produce optimal solutions. But the conception may lead to better feasible points with respect to the set less order relation. In practice, the iteration process is very time-consuming. Since many comparisons can be done in parallel, the descent method can use parallel computation tools. Moreover, adaptive techniques can be applied for the step size control and the refinement of the cluster.

Using nonlinear scalarization results given by Rubinov [23], Tammer (Gerstewitz) [4] and Pascoletti and Serafini [22] in vector optimization and by Hernández and Rodríguez-Marín [7] in set optimization, one then gets other types of min problems which can be solved on a computer. Then even nonconvex set optimization problems should be solvable.

Acknowledgments The author thanks the referees for helpful suggestions and comments and A. Löhne for his advice on Theorem 2.2 and a remark on Proposition 3.1.

References

1. Borwein, J.M.: Multivalued convexity and optimization: a unified approach to inequality and equality constraints. *Math. Program.* **13**, 183–199 (1977)
2. Chiriaev, A., Walster, G.W.: Interval Arithmetic Specification. Technical Report (1998)
3. Dolan, E.D., Lewis, R.M., Torczon, V.: On the local convergence of pattern search. *SIAM J. Optim.* **14**, 567–583 (2003)
4. Gerstewitz (Tammer), C.: Nichtkonvexe Dualität in der Vektoroptimierung. *Wissensch. Zeitschr. TH Leuna-Merseburg* **25**, 357–364 (1983)
5. Gutiérrez, C., Jiménez, B., Miglierina, E., Molho, E.: Scalarization in set optimization with solid and nonsolid ordering cones. *J. Global Optim.* (2014). doi:[10.1007/s10898-014-0179-x](https://doi.org/10.1007/s10898-014-0179-x)

6. Hamel, A.H., Heyde, F.: Duality for set-valued measures of risk. *SIAM J. Financ. Math.* **1**, 66–95 (2010)
7. Hernández, E., Rodríguez-Marín, L.: Nonconvex scalarization in set optimization with set-valued maps. *J. Math. Anal. Appl.* **325**, 1–18 (2007)
8. Jahn, J.: Vectorization in set optimization. *J. Optim. Theory Appl.* (2013). doi:[10.1007/s10957-013-0363-z](https://doi.org/10.1007/s10957-013-0363-z)
9. Jahn, J.: Optimality conditions for set optimization problems with the set less order relation (2014)
10. Jahn, J., Ha, T.X.D.: New order relations in set optimization. *J. Optim. Theory Appl.* **148**, 209–236 (2011)
11. Khoshkhabar-amiranloo, S., Soleimani-damaneh, M.: Scalarization of set-valued optimization problems and variational inequalities in topological vector spaces. *Nonlinear Anal.* **75**, 1429–1440 (2012)
12. Kuroiwa, D.: Natural Criteria of Set-Valued Optimization. Shimane University, Japan (1998)
13. Kuroiwa, D., Tanaka, T., Ha, X.T.D.: On cone convexity of set-valued maps. *Nonlinear Anal.* **30**, 1487–1496 (1997)
14. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.E.: Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Optim.* **9**, 112–147 (1998)
15. Löhne, A.: Private Communication. University of Halle-Wittenberg, Germany (2014)
16. Löhne, A., Schrage, C.: An algorithm to solve polyhedral convex set optimization problems. *Optimization* **62**, 131–141 (2013)
17. McKinnon, K.I.M.: Convergence of the Nelder–Mead simplex method to a nonstationary point. *SIAM J. Optim.* **9**, 148–158 (1998)
18. Nelder, J.A., Mead, R.: A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965)
19. Neukel, N.: Order relations of sets and its application in socio-economics. *Appl. Math. Sci. (Ruse)* **7**, 5711–5739 (2013)
20. Nishnianidze, Z.G.: Fixed points of monotonic multiple-valued operators (in Russian). *Bull. Georgian Acad. Sci.* **114**, 489–491 (1984)
21. Oettli, W.: Optimality conditions for programming problems involving multivalued mappings. In: Korte, B. (ed.) *Modern Applied Mathematics, Optimization and Operations Research*. North-Holland, Amsterdam (1980)
22. Pascoletti, A., Serafini, P.: Scalarizing vector optimization problems. *J. Optim. Theory Appl.* **42**, 499–524 (1984)
23. Rubinov, A.M.: Sublinear operators and their applications. *Russ. Math. Surv.* **32**(4), 115–175 (1977)
24. Sun Microsystems Inc: *Interval Arithmetic Programming Reference*. Sun Microsystems, Palo Alto (2000)
25. Torczon, V.: On the convergence of pattern search algorithms. *SIAM J. Optim.* **7**, 1–25 (1997)
26. Xiao, G., Yu, G., Xiao, H.: Characterizations of scalarization function and set optimization problems with various orders (2012)
27. Young, R.C.: The algebra of many-valued quantities. *Math. Ann.* **104**, 260–290 (1931)