J Optim Theory Appl
DOI 10.1007/s10957-011-9859-6

Stochastic Multiobjective Optimization: Sample Average Approximation and Applications

Jörg Fliege · Huifu Xu

Received: 13 September 2010 / Accepted: 25 April 2011

Abstract We investigate one stage stochastic multiobjective optimization problems where the objectives are the expected values of random functions. Assuming that the closed form of the expected values is difficult to obtain, we apply the well known Sample Average Approximation (SAA) method to solve it. We propose a smoothing infinity norm scalarization approach to solve the SAA problem and analyse the convergence of efficient solution of the SAA problem to the original problem as sample sizes increase. Under some moderate conditions, we show that, with probability approaching one exponentially fast with the increase of sample size, an ϵ-optimal solution to the SAA problem becomes an ϵ-optimal solution to its true counterpart. Moreover, under second order growth conditions, we show that an efficient point of the smoothed problem approximates an efficient solution of the true problem at a linear rate. Finally, we describe some numerical experiments on some stochastic multiobjective optimization problems and report preliminary results.

Keywords Stochastic multiobjective programming · Sample average approximation · Scalarization · Efficient solution · Exponential convergence

1 Introduction

Multiobjective optimization (MOP) problems have become one of the main subject areas in optimization and Operational Research since its foundation by Pareto and

Communicated by Johannes Jahn.

The authors are grateful to the two anonymous referees for their valuable suggestions and comments which have helped improve the presentation of the paper, and to the associate editor for the effective handling of the paper.

J. Fliege (✉) · H. Xu
School of Mathematics, University of Southampton, Southampton, UK
e-mail: J.Fliege@soton.ac.uk

H. Xu
e-mail: H.Xu@soton.ac.uk

Published online: 19 May 2011

Edgeworth for its significant applications in economics, most notably in welfare and utility theory, management and engineering. Like single objective optimization, MOP problems in practice often involve some stochastic data and this motivates one to consider stochastic MOP problems. Earlier research on stochastic multiobjective optimization problems can be found in [1–3]. For more comprehensive references in the area, see [4] and also [5, 6] for recent developments in MOP.

Analogous to single objective optimization, stochastic multiobjective optimization models may take various forms depending on the context of how a decision is made. For instance, if a decision is made after uncertainty is realised, then one may formulate it as a distribution problem. If a decision is made under some probabilistic constraints, then it becomes a chance constrained MOP problem. In this paper, we consider a stochastic MOP problem with deterministic constraints where a decision is made to minimize the expected value of several objective functions that depend not only on the decision variables but also on some random variables. Caballero et al. [4] investigated this model and compared it, in terms of efficient solution sets, with the expected value standard-deviation model where the objective involves not only the expected values but also the standard deviation. The efficient solution sets of the latter model have also been studied by White; see Chap. 8 of [7].

Finding efficient solution sets of a stochastic MOP problem often involves two phases [2]: transforming a stochastic MOP problem into an equivalent deterministic MOP problem and then solving the latter by some MOP approach. This is called a *multiobjective approach* by Abdelaziz [11]. Alternatively, one may change the order by transforming a stochastic MOP problem into a single objective stochastic MOP problem and solving the latter by any stochastic programming approach, and this is known as a *stochastic approach* [11]. Caballero et al. [12] observed that as far as stochastic MOP is concerned, both approaches give rise to efficient solution sets that are identical.

Our focus in this paper is on numerical methods for solving the stochastic MOP problem at hand, i.e. we are interested in approximating the whole set of efficient points, the solution set. One of the main issues to be dealt with is the mathematical expectation in the objective function. Of course, this depends on the availability of information on the random variables and on the properties of the objective functions. If we know the distribution of the random variables, we can integrate out the expected value explicitly. Then, the problem becomes a deterministic MOP and no discretization procedures are required. Throughout this paper, we consider a more interesting case in that the expected values cannot be calculated in a closed form so that we will have to approximate it through some discretization.

A popular numerical method in single objective stochastic programming is the Monte Carlo method where the expected value is approximated by its sample average (sample average approximation, SAA). Over the past years, SAA has been increasingly investigated and recognised as one of the most effective methods for solving single objective stochastic programs; see the discussions in [13–19] and the references therein.

In this paper, we apply this well known method to stochastic MOP. One of our objectives here is to investigate the approximation of the sample averaged problem to the original problem in terms of efficient solutions as the sample size increases. We do so by adopting a well-known max-norm scalarization approach, that is, using a diagonally scaled infinity norm to transform the problem of approximating the sample averaged MOP problem into a family of single-objective optimization problems. To tackle the nonsmoothness resulted from the max-norm, we use a smoothing technique proposed by Peng [20] to smooth the objective function and estimate the error bound of the approximation in terms of a smoothing parameter.

The novel contributions of our paper are as follows:

1. We provide a new family of scalarization techniques for multiobjective problems that, under appropriate assumptions on the objective functions, result in a family of smooth scalar problems whose smoothness can be controlled by a positive real valued smoothing parameter (Lemma 3.1). Consider the set of all solutions of all scalar problems thus constructed. Then, all those solutions are efficient and, in the convex case, the set of all those solutions is a good approximation to the set of efficient points of the original problem as it lies between the set of properly efficient points and the set of efficient points of the original problem (Lemma 3.1 and Theorem 3.1).
2. Consider the set-valued mapping that maps the smoothing parameter onto the set of all solutions of all scalar problems as above. Under appropriate conditions, this mapping is continuous at zero, and the value of the mapping at zero is the set of all solutions of the classical weighted max-scalarization technique (Theorem 3.2 and Corollary 3.3). Moreover, the deviation between the values of the map for positive smoothing parameters and the value of the map at zero shrinks linearly with the smoothing parameter (Theorem 5.1).
3. Consider again the set of all solutions of all scalar problems constructed by the new scalarization technique, but this time with the actual objective function replaced by a sample average. We consider the deviation from this set to the set of almost-optimal points of the original problem. We show that the probability of this deviation being greater than zero converges to zero exponentially fast in the number of samples (Theorem 4.1 and Theorem 5.2).

The rest of this paper is organised as follows. In Sect. 2, we establish the necessary notation, introduce the problem under consideration and introduce some established results on multiobjective optimization and scalarization, to be used later on. Section 3 is concerned with the above mentioned smoothing technique for the max-norm, leading to a new scalarization technique for multiobjective optimization problems. In Sect. 4, we discuss a sample average approximation (SAA) technique for stochastic multiobjective optimization problems and establish that, under suitable conditions, exponential rate of convergence of optimal solutions obtained from solving SAA problems to their true counterparts. Section 5 considers a smoothing technique for a standard nonsmooth scalarization method for multiobjective optimization and show that, under a certain growth condition, the rate of convergence of the optimal solutions of the smoothed problem to their nonsmooth counterparts is linear. Finally, in Sect. 6, we provide some preliminary numerical results.

2 Preliminaries

2.1 Notation

Throughout this paper, we use the following notation. Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}_+$ the set of nonnegative real numbers and $\mathbb{R}_{++}$ the set of positive real numbers. Denote by $\mathbb{R}^n$ the n-dimensional real vector space and let

$$\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \mid \forall i = 1, \dots, n : x_i \geq 0\},$$
$$\mathbb{R}^n_{++} := \{x \in \mathbb{R}^n \mid \forall i = 1, \dots, n : x_i > 0\}$$

be the cones of componentwise nonnegative and strictly positive vectors, respectively. For a set $S \subseteq \mathbb{R}^n$, we denote by $\mathrm{cl}(S)$ its closure.

We denote the standard inner product between two vectors x and y by $\langle x, y\rangle$, while $\|\cdot\|$ denotes the Euclidean norm of a vector and of a compact set of vectors. If D is a compact set of vectors, then

$$\|D\| := \max_{x \in D} \|x\|.$$

Moreover, $d(x, D) := \inf_{y \in D} \|x - y\|$ denotes the distance from point x to set D. For two compact sets D_1 and D_2,

$$\mathbb{D}(D_1, D_2) := \sup_{x \in D_1} d(x, D_2)$$

denotes the deviation from set D_1 to set D_2 (in some references [21] it is also called *excess* of D_1 over D_2). Finally, for a closed convex set $D \subseteq \mathbb{R}^m$ and $x \in D$, we use $\mathcal{N}_C(x)$ to denote the normal cone of D at x, that is,

$$\mathcal{N}_D(x) := \{z \in \mathbb{R}^m : \langle z, y - x\rangle \leq 0, \ \forall\, y \in D\}$$

if $x \in D$ and $\mathcal{N}_D(x) := \emptyset$ if $x \notin D$.

2.2 Problem Statement

Consider a random vector $\xi : \Theta \to \Xi \subset \mathbb{R}^k$ defined on some probability space $(\Theta, \mathcal{F}, P)$, a nonempty set of decision variables $X \in \mathbb{R}^m$ and a vector-valued objective function $f : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^n$, taking as arguments $x \in X$ and $\xi(\theta) \in \mathbb{R}^k$. Each coordinate function f_i $(i = 1, \dots, n)$ corresponds to one objective to be minimized. For $x \in X$, denote by $\mathbb{E}[f(x, \xi(\theta))]$ the expected value of $f(x, \xi(\theta))$ with expectation being taken componentwise. Given some order relation $\lhd$ on the image space $\mathbb{R}^n$ of the objective, we define the stochastic MOP problem to be considered by

$$\min_{\lhd} F(x) := \mathbb{E}\big[f\big(x, \xi(\theta)\big)\big] \quad \text{s.t.} \quad x \in X. \tag{1}$$

We will clarify the precise meaning of the $\min_{\lhd}$ operator after having defined a proper (partial) order of the image space of $F(X)$ below. We call (1) a one stage stochastic

MOP problem as there is only a single level of decisions to be made before the random variable ξ is realised. To simplify notation, we will from now on use ξ to denote either the random vector $\xi(\theta)$ or an element of $\mathbb{R}^k$, depending on the context.

In this paper, we are interested in finding the solution set of the MOP problem with objective function F and a given set of feasible points $X \subseteq \mathbb{R}^m$. As usual in multiobjective optimization, we choose a partial order relation $\lhd$ on the image space $\mathbb{R}^n$ of our objective function (often, $v \lhd w$ if and only if $v_i \leq w_i$ for $i = 1, \dots, n$) and are interested in finding all *efficient* points, i.e. the set

$$\mathcal{E}\big(F(X), \lhd\big) := \big\{F(x) \mid \nexists y \in X : F(y) \lhd F(x) \text{ and } F(y) \neq F(x)\big\} \tag{2}$$

or a good approximation to it, as well as, of course, the corresponding feasible preimages in the decision space X. In contrast to single-objective optimization, it is not sufficient to represent this set by just one point $x \in X$. Instead, the whole set needs to be approximated [8–10].

Let $\xi^1, \dots, \xi^N$ be an identically independent distributed (iid) sample of random variable ξ. Then, one uses the sample average

$$\hat{f}^N(x) := \frac{1}{N} \sum_{j=1}^{N} f\big(x, \xi^j\big) \tag{3}$$

to approximate the expected value $\mathbb{E}[f(x, \xi)]$ and consequently solves the following Sample Average Approximation (SAA) problem

$$\min_{\lhd} \hat{f}^N(x) \quad \text{s.t.} \quad x \in X, \tag{4}$$

by finding the following efficient solution set

$$\mathcal{E}\big(\hat{f}^N(X), \lhd\big) := \big\{\hat{f}^N(x) \mid \nexists y \in X : \hat{f}^N(y) \lhd \hat{f}^N(x) \text{ and } \hat{f}^N(y) \neq \hat{f}^N(x)\big\} \tag{5}$$

or at least an approximation to it. We call (1) the *true problem* which is to find the set $\mathcal{E}(F(X), \lhd)$ (along with corresponding feasible preimages in the decision space X), and (4) the *sample average approximation problem* which is to find the set $\mathcal{E}(\hat{f}^N(X), \lhd)$ (along with corresponding feasible preimages in the decision space X).

2.3 Efficient Points and Scalarizations

Effectively, we consider multiobjective functions

$$F, \hat{f}^N : X \longrightarrow \mathbb{R}^n$$

defined as in (1) and (3) and have to deal with the fact that there is no canonical order in $\mathbb{R}^n$, the image space of our functions, inducing minimal points. We do this as follows. For the rest of this section, let us consider the function F only; everything that follows holds also for the function $\hat{f}^N$. First, define

$$S := F(X)$$

as the image set of our multiobjective problem. Next, let $\lhd$ be an arbitrary order relation on $\mathbb{R}^n$. A vector $v \in \mathbb{R}^n$ is called *minimal* or a *minimizer* w.r.t. $\lhd$ in S iff $v \in S$ and for all $w \in S$ one has $v \lhd w$ or $v = w$. Such a vector v and its corresponding preimage $x = F^{-1}(v)$ would constitute a solution to our multiobjective optimization problem. It is well known that minimal points in the sense defined above usually do not exist. Also, it is not necessarily advisable to use a total order for $\lhd$; see [22] for details.

A weaker concept, the concept of *domination* is therefore needed. A point v *dominates* a point w, if $v \lhd w$ and $v \neq w$ holds. A point v is *nondominated* in S, if $v \in S$ and there does not exist a point $u \in S$ with $u \lhd v$ and $u \neq v$. The concept of nondominated points is the proper generalization of minimal points. Therefore, we define the set of nondominated or *efficient* points of the set S by

$$\mathcal{E}(S, \lhd) := \{v \in S \mid \nexists u \in S : u \lhd v \text{ and } u \neq v\}.$$

The set of solutions of our given multiobjective problem is then $\mathcal{E}(S, \lhd)$, together with the corresponding set of preimages $F^{-1}(\mathcal{E}(S, \lhd))$. Identifying (or approximating) $\mathcal{E}(S, \lhd)$ as well as $F^{-1}(\mathcal{E}(S, \lhd))$ amounts to solving the multiobjective optimization problem at hand.

In what follows, we review briefly standard strategy to compute nondominated elements w.r.t. $\lhd$. First, we need a technical definition, generalizing the concept of ordinary monotonicity.

Definition 2.1 (Monotonicity) Let $S \subseteq \mathbb{R}^n$ be a set and $s : S \to \mathbb{R}$ be a function. The function s is called $\lhd$*-monotonically increasing in* S iff $u \lhd v$ implies $s(u) \leq s(v)$ for all $u, v \in S$. The function s is called *strictly* $\lhd$*-monotonically increasing in* S iff $u \lhd v, u \neq v$ implies $s(u) < s(v)$ for all $u, v \in S$.

Functions monotone with respect to an arbitrary binary relation $\lhd$ are also called *consistent with respect to* $\lhd$; see [7, Chap. 1], or *order-preserving*; see [23, Chap. 7] for an overview. These functions play an important role in multicriteria optimization, as it will be seen in Theorem 2.2 below.

Let $C \subseteq \mathbb{R}^n$ be an arbitrary set and define the order

$$u \lhd_C v \quad :\Longleftrightarrow \quad v - u \in C. \tag{6}$$

The next theorem is well known; see, e.g. [22, 29].

Theorem 2.1 *Let $C \subset \mathbb{R}^n$ be a set and let $\lhd_C$ be the binary relation defined by C as in* (6). *Then, the following statements hold*:

(i) *The relation $\lhd_C$ is translation-invariant in the following sense. For all $u, v, w \in \mathbb{R}^n$ with $u \lhd_C v$ it follows that $(u + w) \lhd_C (v + w)$ holds.*
(ii) *If $0 \in C$ then $\lhd_C$ is reflexive.*
(iii) *If $C + C \subseteq C$ then $\lhd_C$ is transitive.*
(iv) *The set C is a cone if and only if the relation $\lhd_C$ is scale-invariant in the following sense. For all $u, v \in \mathbb{R}^n$ with $u \lhd_C v$ and all $\lambda > 0$ it follows that $\lambda u \lhd_C \lambda v$ holds.*

(v) *If C is a cone containing no lines, i.e. $C \cap -C \subseteq \{0\}$ (such a cone is also called* pointed), *then $\lhd_C$ is anti-symmetric.*
(vi) *The order $\lhd_C$ is total if and only if $C \cup -C = \mathbb{R}^n$.*
(vii) *The set C is closed if and only if the relation $\lhd_C$ is "continuous at 0" in the following sense. For all $v \in \mathbb{R}^n$ and all sequences $(v^{(i)})_{i\in\mathbb{N}}$ in $\mathbb{R}^n$ with $\lim_{i\to+\infty} v^{(i)} = v$ and $0 \lhd_C v^{(i)}$ for all $i \in \mathbb{N}$ it follows that $0 \lhd_C v$ holds.*

According to this result, it is clear that one will usually use a convex pointed cone C to induce the relation $\lhd_C$, unless there are compelling reasons for another choice. The next theorem links binary relations induced by cones and the corresponding efficient points with monotone functions in the sense of Definition 2.1.

Theorem 2.2 *Let $C \subseteq \mathbb{R}^n$ be a cone with $0 \in C$ and $\{0\} \neq C \neq \mathbb{R}^n$, and let $S \subseteq \mathbb{R}^n$ be a set. Let $s : S \to \mathbb{R}$ be a $\lhd_C$-monotone increasing function, and let $v \in S$ be a minimum of s over S. If v is unique or if s is strictly $\lhd_C$-monotone in S, then v is nondominated in S with respect to $\lhd_C$, i.e. $v \in \mathcal{E}(S, \lhd_C)$.*

The proofs can be found in Vogel [24, Chap. 2] or in Göpfert and Nehse [29, Sect. 2.20].

2.4 Linear Scalarizations

The simplest and most widely used $\lhd_C$-monotone functions are the linear forms in $\operatorname{int}(C^*)$, where C^* is the *dual cone* of C defined by

$$C^* = \big\{\omega \mid \forall\, v \in C : \langle \omega, v\rangle \geq 0\big\}.$$

Other nonlinear C-monotone functions have only recently attracted some attention, mainly for numerical reasons [9]. It turns out that in the case of convex cones and sets, only these linear forms need to be considered, at least in theory. More precisely, we need the linear forms from the *quasi-interior* of C^*, i.e. from the set $C^\circ := \{\omega \in \mathbb{R}^n \mid \forall\, v \in C \setminus \{0\} : \langle \omega, v\rangle > 0\}$. With this, it turns out that basically "all" efficient points can be found by minimizing linear functionals $\langle \omega, \cdot\rangle$ over S, as the following theorem shows.

Theorem 2.3 *Let $C, S \subseteq \mathbb{R}^n$ and define*

$$\mathcal{P}(S, C) := \bigcup_{\omega\in C^\circ} \arg\min\big\{\langle \omega, v\rangle \mid v \in S\big\}. \tag{7}$$

Then, the following statements hold:

(i) *Let C be a convex cone with $0 \in C$ and $\{0\} \neq C \neq \mathbb{R}^n$. Then*

$$\mathcal{P}(S, C) \subseteq \mathcal{E}(S, \lhd_C).$$

(ii) *Let C be a closed convex cone with $0 \in C$ such that C contains no lines. Let S be closed and convex. Then*

$$\mathcal{E}(S, \lhd_C) \subseteq \operatorname{cl}\big(\mathcal{P}(S, C)\big).$$

Proofs can be found in various textbooks and original articles; see, e.g. [29, Sect. 2.22], [25, p. 74], or [26]. The first proof of Part (ii) is due to Arrow, Barankin, and Blackwell [27]. Note that the arg min-operator in (7) is understood to work globally, i.e. only global optima are returned.

2.5 Norm Scalarizations

We can also generate all efficient points by minimizing certain norms instead of linear functionals, as the next theorem shows.

Theorem 2.4 *Let $C \subseteq \mathbb{R}^n_+$ be closed convex cone with $C^\circ \neq \emptyset$ and let $S \subseteq \mathbb{R}^n$ be a set such that there exists a $u \in \mathbb{R}^n$ with $S \subseteq u - C$. Then, $v \in \mathcal{E}(S, \lhd_C)$ if and only if there exists a norm $|||\cdot|||$ that is $\lhd_C$-monotone increasing such that v is unique minimum of the function $s(x) := |||x - u|||$.*

For the proof, see Theorem 5.3 of Jahn [28, p. 117]. While the proof is constructive, the use of such norms in a numerical algorithm, however, turns out not to be an easy exercise.

All this leaves us with the question what cone one should use. The most natural choices are, of course, $\mathbb{R}^n_+$ and $\mathbb{R}^n_{++}$. With these particular cones, we define two relations on $\mathbb{R}^n$:

$$v \preceq w \quad :\Longleftrightarrow \quad v \lhd_{\mathbb{R}^n_+} w,$$

$$v \prec w \quad :\Longleftrightarrow \quad v \lhd_{\mathbb{R}^n_{++}} w.$$

Of course, if $v \prec w$, then $v \preceq w$. As above, if $v \preceq w$, we say that v *dominates* the vector w; but if $v \prec w$, we say that v *strictly dominates* the vector w. In multiobjective optimization, we are often concerned with finding the set $\mathcal{E}(S, \preceq)$ or an approximation to it. Less often, we are concerned with finding $\mathcal{E}(S, \prec)$, since this set often contains many more elements than $\mathcal{E}(S, \preceq)$. The set $\mathcal{E}(S, \prec)$ is called the set of *weakly efficient points*.

Note that $\preceq$ is reflexive, while $\prec$ is not. Neither of the two orders is total, but both are translation-invariant, scale-invariant, transitive and anti-symmetric. The cones used to induce these two relations are self-dual, i.e. $C^* = C$ holds, and we also have $(\mathbb{R}^n_+)^\circ = \mathbb{R}^n_{++}$; as such, linear monotone functionals are trivial to find. Note also that, if S is polyhedral, we have

$$\mathcal{E}(S, \preceq) = \mathcal{P}\big(S, \mathbb{R}^n_+\big),$$

see again [29, Sect. 2.22].

Let us now consider particular norms. We start with the max-norm, for which the following result holds.

Theorem 2.5 *Let $S \subseteq \mathbb{R}^n$ be a set such that there exists a $u \in \mathbb{R}^n$ with $S \subseteq \operatorname{int}(u + \mathbb{R}^n_+)$. Then*

$$\mathcal{E}(S, \prec) \subseteq \bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min\big\{ \big\| \operatorname{diag}(\omega)(v - u) \big\|_\infty \,\big|\, v \in S \big\},$$

where the arg min *operator on the right hand side of the equation is understood to return global minima only. Moreover,*

$$\mathcal{E}(S, \prec) = \bigcup_{\omega \in \mathbb{R}^n_{++}} \big\{v \in S \mid v \textit{ is the unique minimizer of } \big\|\operatorname{diag}(\omega)(\cdot - u)\big\|_\infty\big\}.$$

We omit the proof as the results are straightforward.

Note that we do not need convexity assumptions for Theorem 2.5. A standard example [28, p. 298] shows the advantages of using a norm scalarization instead of linear scalarizations in the nonconvex case. Moreover, according to this result, it is not necessary to vary the parameter u. Instead, it is sufficient to choose $u_i := \inf_{s \in S} s_i - 1$ ($i = 1, \dots, n$) or any other suitable lower bound on the minimal values of the F_i available. As such, the assumption of Theorem 2.5 (namely, that such a u exists) relates to the well-posedness of problem (1) in the sense that we assume that each F_i ($i = 1, \dots, n$) is bounded from below.

The max-norm is a nonsmooth function and this may cause some inconvenience in a numerical algorithm for solving the scalarized problem. A natural alternative is to consider the p-norms which are smooth and for which $\|\cdot\|_p \to \|\cdot\|_\infty$ holds for $p \to \infty$.

Let Q be a positive definite diagonal matrix. Then, the function

$$s(v) := \|Qv\|_p$$

is $\preceq$-monotonically increasing on $\mathbb{R}^n_+$ (but not on the whole space $\mathbb{R}^n$). Moreover, a result similar to Theorems 2.3 and 2.5 holds, when employing p-norms instead of the max-norm and varying the parameter p over $[1, +\infty[$; see Theorem 3.4.8 of [25]. Unfortunately, for any *finite* value of p, Remark 3.4.3 of [25] shows that the closure of the efficient set based on a p-norm may not necessarily be contained in $\mathcal{E}(S, \preceq)$, that is, the situation

$$\operatorname{cl}\Bigg(\bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min\big\{\big\|\operatorname{diag}(\omega)(v - u)\big\|_p \,\big|\, v \in S\big\}\Bigg) \subsetneq \mathcal{E}(S, \preceq),$$

can occur, where "cl(S)" denotes the closure of a set. As such, using a fixed p-norm with $p < \infty$ is unsuitable for our problem. This motivates us to consider a smoothing approach for the infinity norm in the next subsection.

3 Smoothing Max-Norm Scalarization

The discussions in the preceding section motivate us to consider a smoothing max-norm scalarization scheme for solving the problem (1). In this section, we explain the technical details about this scheme.

3.1 Reformulations of the True Problem

Let us start by reformulating the true problem as a parameterized single objective minimization problem. The appropriate understanding of "reformulation" should be that an efficient solution point of the true problem is a global minimizer of the reformulated single objective problem under some circumstances and vice versa.

Applying Theorem 2.3, we see that we can find the efficient solution set of the true problem (1), under suitable assumptions, by solving the following single objective optimization problem

$$\min_x \big\langle \omega, F(x) \big\rangle \quad \text{s.t.} \quad x \in X. \tag{8}$$

Let

$$\mathcal{P}\big(F(X), \preceq\big) := \bigcup_{\omega \in (\mathbb{R}^n_+)^\circ} \arg\min\big\{ \langle \omega, v \rangle \mid v \in F(X) \big\}. \tag{9}$$

We will not use this reformulation in our numerical resolution but will need to refer to this problem when we discuss the approximation of efficient sets.

Likewise, by applying Theorem 2.5, we can reformulate (1) as the following max-norm scalarized minimization problem

$$\min_x \big\| \operatorname{diag}(\omega)\big(F(x) - u\big) \big\|_\infty \quad \text{s.t.} \quad x \in X. \tag{10}$$

Let

$$\mathcal{M}\big(F(X), 0\big) := \bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min\big\{ \big\| \operatorname{diag}(\omega)(v - u) \big\|_\infty \,\big|\, v \in F(X) \big\}. \tag{11}$$

(The meaning of the parameter 0 will become clear in Sect. 3.2 below.) The reformulation above means that we can use set $\mathcal{M}(F(X), 0)$ to approximate the efficient solution $\mathcal{E}(F(X), \preceq)$ of the true problem (1) under some appropriate conditions.

Note that there are no convexity assumptions in Theorem 2.5, so using the corresponding approach is more suitable for nonconvex problems or for problems where convexity has not been established. Moreover, we see that, again under suitable assumptions, we can compute the whole set of efficient points (or a very close approximation to it) by varying the parameter $\omega \in \mathbb{R}^n_{++}$; see Theorems 2.3 and 2.5.

3.2 Smoothing Max-Norm Scalarization

We consider now another scalarization technique, which leads to smooth problems while mimicking the behaviour of the max-norm scalarization introduced in Theorem 2.5. Consider the function p defined by

$$p(y, t) := t \log\left(\sum_{i=1}^n e^{y_i / t} \right), \tag{12}$$

where $t > 0$ is a parameter and $p(y, 0) := 0$. This function was introduced by Peng [20] as a smooth approximation of the max function used in numerical methods for solving variational inequality problems.

Using this function, we can introduce a smoothing max-norm scalarization of (10) as follows:

$$\min_x p\big(\operatorname{diag}(\omega)\big(F(x) - u\big), t\big) \quad \text{s.t.} \quad x \in X. \tag{13}$$

We denote the union of all optimal solution sets over all admissible parameters ω by

$$\mathcal{M}\big(F(X), t\big) := \bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min\big\{p\big(\operatorname{diag}(\omega)\big(F(x) - u\big), t\big) \,\big|\, x \in X\big\} \tag{14}$$

and approximate $\mathcal{M}(F(X), 0)$ (and hence the efficient solution set $\mathcal{E}(F(X), \preceq)$ of the true problem (1)) by $\mathcal{M}(F(X), t)$ as t is driven to zero. To this end, we need to investigate the properties of the smoothing function $p(y, t)$. We summarize these in next lemma.

Lemma 3.1 *Let $p(y, t)$ be defined as in (12) and $h(y) := \max_{j=1}^n y_j$. Furthermore, define the set $\mathcal{I}(y) := \{i \in \{1, 2, \dots, n\} \mid h(y) = y_i\}$. Then*

(i) *For fixed $y \in \mathbb{R}^n$, $\lim_{t\to 0} e^{p(y,t)} = \max_{j=1}^n e^{y_j}$ and $\lim_{t\to 0} p(y, t) = h(y)$.*
(ii) *Let $C \subseteq \mathbb{R}^n_+$ be a cone and let $t > 0$ be given. Then, $p(\cdot, t)$ is strictly $\lhd_C$-monotonically increasing.*
(iii) *For any fixed $t > 0$, the function $p(\cdot, t)$ is continuously differentiable and strictly convex in y.*
(iv) *Let $\lambda_i(y, t) := \partial p(y, t)/\partial y_i$, then*

$$\lambda_i(y, t) = \frac{e^{y_i/t}}{\sum_{i=1}^n e^{y_i/t}} \in (0, 1)$$

and $\sum_{i=1}^n \lambda_i(y, t) = 1$.
(v) *Let $t > 0$ be a fixed number and $S \subset \mathbb{R}^n$ be a closed convex set.*[1] *Let a $u \in \mathbb{R}^n$ be given such that $S \subseteq u + \mathbb{R}^n_{++}$. Consider the set-valued mapping*

$$\psi : \omega \mapsto \arg\min_{v \in S} p\big(\operatorname{diag}(\omega)(v - u), t\big).$$

Then, $\psi(\omega)$ is a singleton for all $\omega \in \mathbb{R}^n_{++}$, and the corresponding function is continuous.
(vi) *Let $t > 0$ be fixed and $S \subset \mathbb{R}^n$ be a closed subset in $\mathbb{R}^n$. Assume that there is $u \in \mathbb{R}^n$ such that $S \subseteq u + \mathbb{R}^n_{++}$. Define*

$$\mathcal{M}(S, t) := \bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min_{v \in S} p\big(\operatorname{diag}(\omega)(v - u), t\big). \tag{15}$$

Then, $\mathcal{M}(S, t) \subseteq \mathcal{E}(S, \preceq)$.

[1] Here and in (vi), we use a general set S to keep the statements general, although our focus later on will be on $S = F(X)$.

(vii) *For any* $y \in \mathbb{R}^n_+$ *and* $t > 0$,

$$0 \le p(y,t) - \|y\|_\infty \le t \log n.$$

Proof Part (i). This follows by noting that for $t > 0$, we have $\exp(p(y,t)) = (\sum_{i=1}^n (e^{y_i})^{1/t})^t$. The rest is straightforward.

Part (ii). Let $C \subseteq \mathbb{R}^n_+$ be a cone and let $u, v \in \mathbb{R}^n$ be given with $u \neq v$ and $u \lhd_C v$, i.e. $v - u \in C \subseteq \mathbb{R}^n_+$. As a consequence, $u_i \le v_i$ and therefore $e^{u_i/t} \le e^{v_i/t}$ for $i = 1, \dots, n$, but $e^{u_j/t} < e^{v_j/t}$ for one $j \in \{1, \dots, n\}$. From this, the result follows.

Parts (iii) and (iv). See [20, Lemma 2.1].

Part (v). This follows from the strict convexity of $p(\cdot, t)$ and Proposition 4.32 of [31].

Part (vi). Let $v \in \mathcal{M}(S,t)$, i.e. v minimizes $q(v,t) := p(\mathrm{diag}(\omega)(v-u), t)$ over S for some given $u \in \mathbb{R}^n$, $\omega \in \mathbb{R}^n_{++}$ and $t > 0$. According to parts (ii) and (iii), $p(\cdot, t)$ is strictly $\preceq$-monotonically increasing and strictly convex, and therefore $q(\cdot, t)$ is strictly $\preceq$-monotonically increasing and strictly convex. Through Theorem 2.2, $w \in \mathcal{E}(S, \preceq)$. The conclusion follows.

Part (vii). The assertion was made by Peng [20, equation (2.2)] for $p(y,t) - \max(y_i)$ where the components of y are not restricted to be nonnegative. Here we provide a proof for completeness and for the fact that the conclusion may not hold when some components of y are negative.

The first inequality is straightforward. We only prove the second. Let I_0 denote the index set such that

$$\|y\|_\infty = y_{i_0}, \quad i_0 \in I_0.$$

Then

$$p(y,t) = \log\left(e^{y_{i_0}}\left(|I_0| + \sum_{i \neq i_0} e^{\frac{y_i - y_{i_0}}{t}}\right)^t\right)$$

where $|I_0|$ denotes the cardinality of I_0. Since $y_i < y_{i_0}$, then $e^{\frac{y_i - y_{i_0}}{t}} < 1$, which implies

$$p(y,t) \le \|y\|_\infty + t \log n,$$

where n is the dimension of vector y. Equality holds when $|I_0| = n$. □

Note also that part (vi) of the preceding lemma holds for *arbitrary* $t > 0$. Especially, we have the following corollary in the case when S is convex.

Theorem 3.1 *Let* $t > 0$ *be fixed and* $S \subset \mathbb{R}^n$ *be a closed convex set. Let there be given an* $u \in \mathbb{R}^n$ *such that* $S \subseteq u + \mathbb{R}^n_{++}$. *Then*

$$\mathcal{P}(S, \preceq) \subseteq \mathcal{M}(S,t) \subseteq \mathcal{E}(S, \preceq) \subseteq \mathrm{cl}\big(\mathcal{P}(S, \preceq)\big)$$

where $\mathcal{P}(S,\preceq) := \mathcal{P}(S,\mathbb{R}^n_+)$. *Moreover, if the set* $\Omega \subseteq \mathbb{R}^n_{++}$ *is a dense subset of* $\mathbb{R}^n_{++}$, *then*

$$\bigcup_{\omega\in\Omega} \arg\min_{v\in S} p\big(\operatorname{diag}(\omega)(v-u),t\big)$$

is dense in $\mathcal{E}(S,\preceq)$.

Proof Let $w \in \mathcal{P}(S,\preceq)$. Then, there exists an $\omega \in \mathbb{R}^n_{++}$ such that $-\omega \in \mathcal{N}_S(w)$. For $t>0$, define $\varrho \in \mathbb{R}^n_{++}$ implicitly as the unique solution to the equations $\varrho_i := \omega_i / e^{\varrho_i(w_i-u_i)/t}$ for $i=1,\dots,n$. Consider the function $q(v) := p(\operatorname{diag}(\varrho)(v-u),t)$ and note that

$$\frac{\partial q(w)}{\partial w_i} = \frac{\varrho_i e^{\varrho_i(w_i-u_i)/t}}{\sum_{j=1}^n e^{\varrho_j(w_j-u_j)/t}}.$$

With this, $-\nabla q(w) = -\omega / \sum_{j=1}^n e^{\varrho_j(w_j-u_j)/t} \in \mathcal{N}_S(w)$, and therefore w is the unique minimizer of q over S and as such an element of $\mathcal{M}(S,t)$. The density result follows immediately from Lemma 3.1(v). □

According to this result, scalarizing a multiobjective problem with image set S has, under the stated assumptions, two advantages:

1. It is *not* necessary to use small parameter values $t\approx 0$ when solving convex multiobjective optimization problems with this scalarization technique. In this case, each member $w\in\mathcal{M}(S,t)$ is the unique minimizer of a strictly convex objective function of the form $p(\operatorname{diag}(\omega)(\cdot-u),t)$.
2. A sufficiently good discretization of the parameter space $\mathbb{R}^n_{++}$ (or of a base generating this cone) by parameter vectors ω will result in a good approximation of $\mathcal{E}(S,\preceq)$, a much sought-after situation in multiobjective optimization [9].

For the general case, we have the following result.

Theorem 3.2 *Extend the definition of the mapping* $\mathcal{M}$ *by setting*

$$\mathcal{M}(S,0) := \bigcup_{\omega\in\mathbb{R}^n_{++}} \arg\min\big\{\big\|\operatorname{diag}(\omega)(v-u)\big\|_\infty \,\big|\, v\in S\big\}. \tag{16}$$

Then the mapping $\mathcal{M}(S,\cdot)$ *is upper semicontinuous at* $t=0$.

Furthermore, let $S\subseteq u+\mathbb{R}^n_{++}$ *be compact. For each* $\omega\in\mathbb{R}^n_{++}$ *and* $t\ge 0$ *sufficiently small (including* $t=0$ *in which case we refer to the max-norm), if* $\|\operatorname{diag}(\omega)(\cdot-u)\|_{1/t}$ *has a unique minimizer in* S, *then the mapping* $\mathcal{M}(S,\cdot)$ *is continuous at* $t=0$.

Proof Let $u\in\mathbb{R}^n$ be given with $S\subseteq u+\mathbb{R}^n_{++}$. Without loss of generality, $u=0$. Consider the coordinate transformation $z=g(y):=(e^{y_1},\dots,e^{y_n})$ $(y\in\mathbb{R}^n)$. Due to [24, Satz 2.1], we have $\mathcal{E}(S,\lhd) = g^{-1}(\mathcal{E}(g(S),\lhd))$ for all orders $\lhd$ with $u\lhd v \Leftrightarrow$

$g(u) \lhd g(v)$. The latter is clearly the case for $\lhd = \preceq$ and $\lhd = \prec$, and we see that we only need to consider problems in the z-coordinates. Let $Z = g(S)$,

$$\mathcal{G}(Z,t) := \bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min\left\{ \left\| \operatorname{diag}(\omega) z \right\|_{1/t} \,\middle|\, z \in Z \right\}$$

for $t > 0$ and

$$\mathcal{G}(Z,0) := \bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min\left\{ \left\| \operatorname{diag}(\omega) z \right\|_{\infty} \,\middle|\, z \in Z \right\}.$$

By exploiting the monotonicity of the logarithm, we see that

$$g^{-1}\big(\mathcal{G}(Z,t)\big) = \mathcal{M}(S,t).$$

Therefore, it suffices to consider the semicontinuity of the mapping $\mathcal{G}$.

Let $\omega \in \mathbb{R}^n_{++}$ be given. Observe that for any $z \in Z$, $\|\operatorname{diag}(\omega)z\|_{1/t}$ converges to $\|\operatorname{diag}(\omega)z\|_\infty$ which implies pointwise convergence on Z. Moreover, $\|\operatorname{diag}(\omega)z\|_{1/t}$ is globally Lipschitz continuous, which implies that the function is equi-continuous on Z. Through [30, Theorem 7.10], the two properties imply that $\|\operatorname{diag}(\omega)\cdot\|_{1/t}$ epiconverges to $\|\operatorname{diag}(\omega)\cdot\|_\infty$ on $\mathbb{R}^n$ and through [31, Proposition 4.6] the upper semicontinuity of its solution set as t tends to 0.

To prove the continuity under the additional conditions, it suffices to show lower semicontinuity of $\mathcal{G}(Z,\cdot)$ at $t = 0$. We proceed as follows. Let $\omega \in \mathbb{R}^n_{++}$ and $t > 0$ be given, let $z(\omega,t)$ be the unique minimizer of $\|\operatorname{diag}(\omega)\cdot\|_{1/t}$ in Z and let $v(\omega,t) := \|\operatorname{diag}(\omega)z(\omega,t)\|_{1/t}$. For $z \in Z$, we introduce a domination structure $D(z,\omega,t)$ in the sense of [32] by

$$D(z,\omega,t) := -z + \left\{ \hat{z} \in \mathbb{R}^n \mid v(\omega,t) \geq \left\| \operatorname{diag}(\omega)\hat{z} \right\|_{1/t} \right\}.$$

This domination structure induces a binary relation $\lhd_t$: for all $z, \hat{z} \in \mathbb{R}^n$ we have $z \lhd_t \hat{z}$ if and only if $\hat{z} \in z + D(z,\omega,t)$. By definition, the set of efficient points of Z with respect to this domination structure is

$$\mathcal{E}(Z,\lhd_t) = \left\{ \hat{z} \in Z \mid \nexists z \in Z : z \neq \hat{z} \text{ and } \hat{z} \in z + D(z,\omega,t) \right\}.$$

We clearly have $\mathcal{E}(Z,\lhd_t) = \{z(\omega,t)\}$ and

$$Z \subseteq \left\{ z \in \mathbb{R}^n \,\middle|\, \left\| \operatorname{diag}(\omega) z(\omega,t) \right\|_{1/t} \leq \left\| \operatorname{diag}(\omega) z \right\|_{1/t} \right\},$$

i.e. the set $\mathcal{E}(Z,\lhd_t)$ is externally stable in the sense of Definition 3.2.6 (p. 59) of [25]. Moreover, $D(\cdot,\omega,\cdot)$ is an upper semicontinuous mapping. As a consequence, the mapping $t \mapsto \mathcal{E}(Z,\lhd_t)$ is lower semicontinuous for all $\omega \in \mathbb{R}^n_{++}$; see Theorem 4.3.2 (p. 112) of [25]. From this, the result follows. □

Note that it suffices to assume that S is convex and closed in order to ensure that the functions $\|\operatorname{diag}(\omega)(\cdot - u)\|_{1/t}$ have unique minimizers in S.

In case $\mathcal{E}(S,\preceq)=\mathcal{E}(S,\prec)$ holds, upper semicontinuity can also be shown by invoking Theorem 2.5 as well as Part 6 of Lemma 3.1.

Note that the function p essentially introduces a coordinate transformation in $\mathbb{R}^n$ first before mapping vectors to real numbers: each single coordinate is transformed with the help of the strictly monotone and strictly convex functions $g_i : v_i \mapsto e^{\omega_i(v_i-u_i)/t}$ before the actual scalarization takes place, i.e. before these new coordinates are summed up. Accordingly, each face of $g(S)$ has at most one point with $\mathcal{E}(g(S),\preceq)$ in common, a highly desirable property for numerical algorithms; see Theorem 3.2 of [9] and the accompanying discussion.

Corollary 3.1 *Let X be closed and convex and let the function $F(x) := \mathbb{E}[f(x,\xi)]$ be convex and continuous on X. Furthermore, let $t>0$ be fixed and $u\in\mathbb{R}^n$ be given with $u_i < F_i(x)$ for all $x\in X$ $(i=1,\dots,n)$. Then*

$$\mathcal{P}\big(F(X),\preceq\big)\subseteq\mathcal{M}\big(F(X),t\big)\subseteq\mathcal{E}\big(F(X),\preceq\big)\subseteq\operatorname{cl}\big(\mathcal{P}\big(F(X),\preceq\big)\big)$$

where $\mathcal{P}(F(X),\preceq)$ is defined as in (9). Moreover, if the set $\Omega\subseteq\mathbb{R}^n_{++}$ is a dense subset of $\mathbb{R}^n_{++}$, then

$$\bigcup_{\omega\in\Omega}\arg\min_{F(x)\in F(X)} p\big(\operatorname{diag}(\omega)\big(F(x)-u\big),t\big)$$

is dense in $\mathcal{E}(F(X),\preceq)$.

Corollary 3.2 *Corollary 3.1 holds if F is replaced by $\hat{f}^N$ throughout.*

Clearly, the sets $\mathcal{P}(F(X),\preceq)$ and $\operatorname{cl}(\mathcal{P}(F(X),\preceq))$ are indistinguishable from each other in floating-point arithmetic. As such, under the conditions of the last corollaries, the sets $\mathcal{M}(F(X),t)$ and $\mathcal{E}(F(X),\preceq)$ are indistinguishable from each other in floating-point arithmetic. In other words, $\mathcal{M}(F(X),t)$ is a perfect approximation of the set that we want to approximate, and it is sufficient to use a sufficiently dense discretization of the parameter space.

Corollary 3.3 *Let $\mathcal{M}(F(X),t)$ and $\mathcal{M}(F(X),0)$ be defined by (14) and (11), respectively. Then $\mathcal{M}(F(X),t)$ is upper semicontinuous at $t=0$. Furthermore, let X be compact and F be continuous, and let $u\in\mathbb{R}^n$ be given with $u_i<F_i(x)$ for all $x\in X$ $(i=1,\dots,n)$. For each $\omega\in\mathbb{R}^n_{++}$ and each $t\geq 0$ sufficiently small (including $t=0$), assume that the problem*

$$\min \big\|F(x)-u\big\|_{1/t}\quad s.t.\quad x\in X \tag{17}$$

has a unique solution. Then, the mapping $\mathcal{M}(F(X),t)$ is continuous at $t=0$.

Corollary 3.4 *Corollary 3.3 holds if F is replaced by $\hat{f}^N$ throughout.*

4 Sample Average Approximation

In this section, we move on to consider the following problem

$$\min \left\| \operatorname{diag}(\omega)\big(\hat{f}^N(x) - u\big)\right\|_\infty \quad \text{s.t.} \quad x \in X, \tag{18}$$

where $\omega \in \mathbb{R}^n_{++}$ is arbitrary and we assume that $u \in \mathbb{R}^n$ has been chosen in such a way that $u < f(x,\xi)$ for all $(x,\xi) \in X \times \Xi$ holds. Let

$$\mathcal{M}\big(\hat{f}^N(X), 0\big) := \bigcup_{\omega \in \mathbb{R}^n_{++}} \arg\min\big\{\left\|\operatorname{diag}(\omega)(v-u)\right\|_\infty \,\big|\, v \in \hat{f}^N(X)\big\}. \tag{19}$$

Of course, it is not necessary in the definition of $\mathcal{M}(\hat{f}^N(X), 0)$ to take the union over all $\omega \in \mathbb{R}^n_{++}$. It suffices to take the union over all $\omega \in \mathbb{R}^n_{++}$ with $\|\omega\| \le 1$, and we will do so if this simplifies the derivation.

Obviously, (18) is the single-objective max-norm scalarization of the sample average approximation problem (4). In this section, we investigate approximation of $\mathcal{M}(F(X), 0)$ by $\mathcal{M}(\hat{f}^N(X), 0)$ as the sample size N increases.

Assume now that we can solve (18) and obtain a global optimal solution $x^N(\omega)$. First, we investigate the convergence of $\{x^N(\omega)\}$ as the sample size N increases. For this, we need the following intermediate result.

Lemma 4.1 *Consider a general constrained minimization problem*

$$\min \phi(x) \quad s.t. \quad x \in X \tag{20}$$

where $\phi : \mathbb{R}^m \to \mathbb{R}$ is continuous and $X \subseteq \mathbb{R}^m$ is closed, and a perturbed program

$$\min \psi(x) \quad s.t. \quad x \in X \tag{21}$$

where $\psi : \mathbb{R}^m \to \mathbb{R}$ is continuous and $|\psi(x) - \phi(x)| \le \delta$, $\forall x \in X$. Let x^* be a global minimizer of ϕ over X, and $\tilde{x}$ a global minimizer of ψ over X. Then

(i) $|\phi(x^*) - \psi(\tilde{x})| \le \sup_{x\in X} |\psi(x) - \phi(x)| \le \delta$.
(ii) *A global minimizer of ψ is a δ-global minimizer of ϕ and vice versa, that is, $|\phi(x^*) - \phi(\tilde{x})| \le 2\delta$ and $|\psi(x^*) - \psi(\tilde{x})| \le 2\delta$.*

The result is well-known, see, for instance, [17]. Note that, strictly speaking, x^* is a 2δ-global minimizer of ψ. However, by convention, we call it a δ-global minimizer.

Let $f(x,\xi)$ be defined as in (1). We denote by

$$M_x(t) := \mathbb{E}\big\{e^{t\|f(x,\xi) - \mathbb{E}[f(x,\xi)]\|}\big\}$$

the moment generating function of the random variable $\|f(x,\xi) - \mathbb{E}[f(x,\xi)]\|$. Let us make the following assumptions.

Assumption 4.1

(a) For every $x \in X$, the moment generating function $M_x(t)$ is finite valued for all t in a neighbourhood of zero.
(b) There exists an integrable function $\kappa : \Xi \to \mathbb{R}_+$ and a constant $\gamma > 0$ such that

$$\left\| F(y, \xi) - F(x, \xi) \right\| \le \kappa(\xi) \|y - x\|^{\gamma} \tag{22}$$

for all $\xi \in \Xi$ and all $x, y \in X$.
(c) The moment generating function $M_\kappa(t)$ of $\kappa(\xi)$ is finite valued for all t in a neighbourhood of zero.

Assumptions 4.1(a) and (c) mean that the probability distributions of the random variable $\|f(x, \xi)\|$ and the random variable $\kappa(\xi)$ die exponentially fast in the tails. In particular, it holds if the random variables have a distribution supported on a bounded subset of $\mathbb{R}$.

Lemma 4.2 *Let Assumption 4.1 hold and X be a compact set. Then for any $\epsilon > 0$ there exist positive constants $c(\epsilon)$ and $\beta(\varepsilon)$, independent of N, such that*

$$\mathrm{Prob}\left\{ \sup_{x \in X} \left\| \hat{f}^N(x) - \mathbb{E}\big[f(x, \xi)\big] \right\| \ge \epsilon \right\} \le c(\epsilon) e^{-N\beta(\epsilon)} \tag{23}$$

for N sufficiently large.

Proof The lemma is a simple generalization of [33, Theorem 5.1] where f is a real valued function. The conclusion follows by applying [33, Theorem 5.1] to each component function of $\hat{f}^N$ here. □

Theorem 4.1 *Let Assumption 4.1 hold and X be a compact set. Let $\mathcal{M}(\hat{f}^N(X), 0)$ be defined as in (19). Let $\epsilon > 0$ be a fixed scalar and $v(\omega)$ denote the optimal value of minimization problem (10). Define*

$$\mathcal{M}^{\epsilon}\big(F(X), 0\big) := \bigcup_{\substack{\omega \in \mathbb{R}^n_{++} \\ \|\omega\| \le 1}} \Big\{ x \in X : \big\| \mathrm{diag}(\omega)\big(F(x) - u\big) \big\|_\infty \le v(\omega) + 2\epsilon \Big\}.$$

Then

(i) *With probability approaching one exponentially fast with the increase of sample size N, the optimal value of (18) converges to that of (10) uniformly with respect to ω;*
(ii) *For any $\epsilon > 0$, there exist constants $\hat{c} > 0$ and $\hat{\beta} > 0$ (dependent of ϵ) such that*

$$\mathrm{Prob}\Big(\mathbb{D}\big(\mathcal{M}\big(\hat{f}^N(X), 0\big), \mathcal{M}^{\epsilon}\big(F(X), 0\big)\big) > 0\Big) \le \hat{c} e^{-\hat{\beta} N} \tag{24}$$

for N sufficiently large.

Proof Observe first that, by the marginal map theorem [34, Theorem 8.2.11], the set $\mathcal{M}(\hat{f}^N(X),0)$ is measurable. For fixed $\omega \in \mathbb{R}^n_{++}$, let $x^*(\omega)$ denote a global optimal solution of (10) and $x^N(\omega)$ denote an optimal solution of (18). Without loss of generality, assume $\|\omega\| = 1$.

Part (i). By definition, $v(\omega) = \|\mathrm{diag}(\omega)(F(x^*(\omega)) - u)\|_\infty$, and by Lemma 4.1

$$\begin{aligned} &\left|v(\omega) - \left\|\mathrm{diag}(\omega)\left(\hat{f}^N\left(x^N(\omega)\right) - u\right)\right\|_\infty\right| \\ &\quad \le \sup_{x\in X}\left|\left\|\mathrm{diag}(\omega)\left(F(x) - u\right)\right\|_\infty - \left\|\mathrm{diag}(\omega)\left(\hat{f}^N(x) - u\right)\right\|_\infty\right| \\ &\quad \le \|\omega\| \sup_{x\in X}\left\|\hat{f}^N(x) - F(x)\right\| \\ &\quad = \sup_{x\in X}\left\|\hat{f}^N(x) - F(x)\right\|. \end{aligned} \tag{25}$$

Under Assumption 4.1, it follows from Lemma 4.2 and the discussion above that (23) holds. The conclusion follows.

Part (ii). Observe that

$$\mathbb{D}\left(\mathcal{M}\left(\hat{f}^N(X),0\right), \mathcal{M}^\epsilon\left(F(X),0\right)\right) > 0 \tag{26}$$

implies

$$\sup_{\substack{\omega\in\mathbb{R}^n_{++}\\ \|\omega\|\le 1}} \sup_{x\in X}\left|\left\|\mathrm{diag}(\omega)\left(F(x) - u\right)\right\|_\infty - \left\|\mathrm{diag}(\omega)\left(\hat{f}^N(x) - u\right)\right\|_\infty\right| \ge \epsilon. \tag{27}$$

To see this, let $\omega \in \mathbb{R}^n_{++}$, $\|\omega\| = 1$ and $x^N(\omega) \in \mathcal{M}(\hat{f}^N(X),0)$ be such that $x^N(\omega)$ is a solution to $\min_{x\in X} \|\mathrm{diag}(\omega)(\hat{f}^N(x) - u)\|_\infty$ and

$$\mathbb{D}\left(\mathcal{M}\left(\hat{f}^N(X),0\right), \mathcal{M}^\epsilon\left(F(X),0\right)\right) = d\left(x^N(\omega), \mathcal{M}^\epsilon\left(F(X),0\right)\right) > 0.$$

The existence of $x^N(\omega)$ is guaranteed as $\mathcal{M}(\hat{f}^N(X),0)$ is a nonempty compact set. Obviously, (26) implies that $x^N(\omega) \notin \mathcal{M}^\epsilon(F(X),0)$ and, through the definition of $\mathcal{M}^\epsilon(F(X),0)$,

$$\left\|\mathrm{diag}(\omega)\left(F\left(x^N(\omega)\right) - u\right)\right\|_\infty > v(\omega) + 2\epsilon.$$

In view of Lemma 4.1(ii), this implies

$$\sup_{x\in X}\left|\left\|\mathrm{diag}(\omega)\left(F(x) - u\right)\right\|_\infty - \left\|\mathrm{diag}(\omega)\left(\hat{f}^N(x) - u\right)\right\|_\infty\right| \ge \epsilon,$$

and hence (27). Using the relationship between (26) and (27), we have

$$\begin{aligned}
&\operatorname{Prob}\big\{\mathbb{D}\big(\mathcal{M}\big(\hat{f}^N(X),0\big),\mathcal{M}^{\epsilon}\big(F(X),0\big)\big)>0\big\}\\
&\quad\le \operatorname{Prob}\Big\{\sup_{\substack{\omega\in\mathbb{R}^n_{++},\\ \|\omega\|\le 1}}\sup_{x\in X}\big|\big\|\operatorname{diag}(\omega)\big(F(x)-u\big)\big\|_\infty-\big\|\operatorname{diag}(\omega)\big(\hat{f}^N(x)-u\big)\big\|_\infty\big|\ge\epsilon\Big\}\\
&\quad\le \operatorname{Prob}\Big\{\sup_{x\in X}\big\|\hat{f}^N(x)-F(x)\big\|\ge\epsilon\Big\},
\end{aligned}$$

where the last inequality is due to (25). The rest follows from Lemma 4.2. □

5 Smoothing Approximation

The problem under consideration in (18) is a nonsmooth minimization problem. Algorithms which can be used to solve such a problem include, e.g. the well-known bundle methods [35, 36], particularly when $f(x,\xi)$ is convex in x. Here, however, we will consider a smoothing method based on (12), for the following reasons:

1. The nonsmoothness in (18) is only caused by the max-operator of the norm, so it should be advantageous to exploit this particular structure.
2. The function p from (12) is strictly $\lhd_C$-monotonically increasing for a large set of cones C (see Part 3 of Lemma 3.1) and its usage comes as such naturally in a multiobjective framework.
3. In the convex case, Part 6 of Lemma 3.1 ensures us that it is not necessary to use large smoothing parameter that would have a detrimental effect on the numerical stability of the solution process.

Consider the same parameters $\omega\in\mathbb{R}^n_{++}$, $u\in\mathbb{R}^n$ as used in (18). Using the smoothing function introduced in (12), we consider a smooth approximation of problem (18), namely

$$\min p\big(\operatorname{diag}(\omega)\big(\hat{f}^N(x)-u\big),t\big)\quad \text{s.t.}\quad x\in X, \tag{28}$$

and its true counterpart defined in (13).

From Lemma 3.1, both objective functions for these two problems are continuously differentiable for $t>0$ if $F, f(\cdot,\xi)\in C^1$ for all ξ. Moreover, they are convex in x if $f(\cdot,\xi)$ is convex for all ξ. In what follows, we investigate how the global optimal solution of (13) approximates that of (10) as t tends to 0.

Lemma 5.1 *Consider problems* (20) *and* (21) *as defined in Lemma* 4.1. *Let* X^*_ϕ *denote the set of optimal solutions to* (20) *and* X^*_ψ *the set of optimal solutions to* (21). *Then for any* $\epsilon>0$, *there exists a* $\delta>0$ (*depending on* ϵ) *such that*

$$\mathbb{D}\big(X^*_\psi,X^*_\phi\big)\le\epsilon, \tag{29}$$

when

$$\sup_{x\in X}\big|\psi(x)-\phi(x)\big|\le\delta.$$

If, in addition, there exists $\nu > 0$ *such that*

$$\phi(x) \geq \min_{x\in X} \phi(x) + \alpha d\big(x, X_\phi^*\big)^\nu, \quad \textit{for all } x \in X, \tag{30}$$

then

$$\mathbb{D}\big(X_\psi^*, X_\phi^*\big) \leq \left(\frac{3}{\alpha} \sup_{x\in X} \big|\psi(x) - \phi(x)\big|\right)^{\frac{1}{\nu}}. \tag{31}$$

Proof The result is a minor extension of [37, Lemma 3.2] which deals with the case when X_ϕ^* is a singleton. Here we provide a proof for completeness. Let ϵ be a fixed small positive number and ϕ^* the optimal value of (20). Define

$$R(\epsilon) := \inf_{\{x\in X, d(x, X_\phi^*)\geq\epsilon\}} \phi(x) - \phi^*.$$

Then $R(\epsilon) > 0$. Let $\delta := R(\epsilon)/3$ and ψ be such that $\sup_{x\in X} |\psi(x) - \phi(x)| \leq \delta$. Then for any $x \in X$ with $d(x, X_\phi^*) \geq \epsilon$ and any fixed $x^* \in X_\phi^*$,

$$\psi(x) - \psi(x^*) \geq \phi(x) - \phi(x^*) - 2\delta \geq R(\epsilon)/3 > 0,$$

which implies that x is not an optimal solution to (21). This is equivalent to $d(x, X_\phi^*) < \epsilon$ for all $x \in X_\psi^*$, that is, $\mathbb{D}(X_\psi^*, X_\phi^*) \leq \epsilon$.

Let us now consider the case when condition (30) holds. In such a case, it is easy to derive that $R(\epsilon) = \alpha\epsilon^\nu$. Therefore, (31) follows by setting

$$\epsilon := \left(\frac{3}{\alpha} \sup_{x\in X} \big|\psi(x) - \phi(x)\big|\right)^{\frac{1}{\nu}}$$

in the first part of the proof. □

Note that in the case when $\nu = 2$, (30) is known as a second order growth condition. The terminology was introduced by Shapiro [38] for the stability analysis of stochastic programming and has been widely used afterwards; see [31, 39, 40] and the references therein.

We are now ready to state the main result of this section, Theorem 5.1, which describes the deviation of a solution to the smoothed problem (13) from the solution set of (10).

Theorem 5.1 *Let* ω *be fixed and* $X^*(\omega)$ *and* $X(\omega, t)$ *denote the set of optimal solutions of* (10) *and* (13), *respectively. Let* $F(x) := \mathbb{E}[f(x, \xi)]$ *and* $\mathcal{I}(x) := \{i \mid \omega_i F_i(x) = \max_j \omega_j F_j(x)\}$. *Assume*: (a) $f_i(x, \xi)$, $i = 1, \ldots, n$, *is convex in* x, (b) $\max_i(\omega_i F_i(x))$ *satisfies the growth condition, that is, there exist positive constants* α *and* ν *independent of* ω *such that*

$$\max_i\big(\omega_i F_i(x)\big) \geq \min_{x\in X} \max_i\big(\omega_i F_i(x)\big) + \alpha d\big(x, X^*(\omega)\big)^\nu, \quad \textit{for all } x \in X. \tag{32}$$

Then we have the following:

(i) *For any* $t > 0$,

$$\mathbb{D}\big(X(\omega,t), X^*(\omega)\big) \le \left(\frac{3}{\alpha} t \log n\right)^{\frac{1}{\nu}}. \tag{33}$$

(ii) *Let* $\mathcal{M}(F(X),0)$ *be defined as in* (11) *and* $\mathcal{M}(F(X),t)$ *in* (14). *Then*

$$\mathbb{D}\big(\mathcal{M}\big(F(X),t\big), \mathcal{M}\big(F(X),0\big)\big) \le \left(\frac{3}{\alpha} t \log n\right)^{\frac{1}{\nu}}.$$

Proof Part (ii) follows from Part (i) and the definition of sets $\mathcal{M}(F(X),0)$ and $\mathcal{M}(F(X),t)$. So we only prove Part (i).

Under the growth condition (32), it follows from Lemma 5.1 that

$$\begin{aligned}
&\mathbb{D}\big(X(\omega,t), X^*(\omega)\big) \\
&\quad \le \left(\frac{3}{\alpha} \sup_{x\in X} \big|p\big(\mathrm{diag}(\omega)\big(F(x)-u\big),t\big) - \big\|\mathrm{diag}(\omega)\big(F(x)-u\big)\big\|_\infty\big|\right)^{\frac{1}{\nu}}. \qquad (34)
\end{aligned}$$

By Lemma 3.1(vii),

$$\big|p\big(\mathrm{diag}(\omega)\big(F(x)-u\big),t\big) - \big\|\mathrm{diag}(\omega)\big(F(x)-u\big)\big\|_\infty\big| \le t\log n.$$

This gives (33). The proof is complete. □

Finally, we study the approximation of (28) to (13) as sample size N increases. We have the following.

Theorem 5.2 *Let* $t > 0$ *and*

$$\mathcal{M}\big(\hat{f}^N(X),t\big) := \bigcup_{\omega\in\mathbb{R}^n_{++}} \arg\min\big\{p\big(\mathrm{diag}(\omega)(v-u),t\big)\,\big|\, v\in \hat{f}^N(X)\big\} \tag{35}$$

and

$$\mathcal{M}^\epsilon\big(F(X),t\big) := \bigcup_{\omega\in\mathbb{R}^n_{++}} \big\{x\in X : p\big(\mathrm{diag}(\omega)\big(F(x)-u\big),t\big) \le v(\omega,t) + 2\epsilon\big\},$$

where ϵ *is a fixed positive scalar and* $v(\omega,t)$ *denotes the optimal value of minimization problem* (13). *Assume the conditions of Lemma* 4.2. *Then with probability approaching one exponentially fast with the increase of sample size* N, *the optimal value of* (28) *converges to that of* (13) *uniformly with respect to* ω, *that is, for any* $\epsilon > 0$, *there exist constants* $\hat{c}(t) > 0$ *and* $\hat{\beta}(t) > 0$ (*dependent of* ϵ) *such that*

$$\mathrm{Prob}\big(\mathbb{D}\big(\mathcal{M}\big(\hat{f}^N(X),t\big), \mathcal{M}^\epsilon\big(F(X),t\big)\big) > 0\big) \le \hat{c}(t)e^{-\hat{\beta}(t)N} \tag{36}$$

for N *sufficiently large.*

The proof is similar to that of Theorem 4.1 by replacing $\|\cdot\|_\infty$ with $p(\cdot, t)$ and using Lemma 3.1. We omit the details. Note that if we want to obtain a rate of convergence of the solution set of (28) to that of (13), we need some growth condition of $p(\mathrm{diag}(\omega)F(x), t)$ for fixed t and it is unclear whether this is implied by the growth condition (32). The analysis is similar to that of Theorem 5.1 by exploiting Lemma 5.1. Again, we omit the details.

6 Numerical Results

In this section, we illustrate the characteristics of the model (1) by discussing several potential applications to it. We describe the corresponding numerical experiments on these problems and report some preliminary numerical test results. Optimization of the sampled average functions took place within MATLAB Version 7.2.0.294 (R2006a) using routine `ucSolve` with standard parameter settings from the TOMLAB Version 6.0 package.

6.1 A Multiobjective Stochastic Location Problem

For $a \in \mathbb{R}^m$ and a random vector $\xi \in \mathbb{R}^m$, consider the unconstrained bi-objective optimization problem with decision variables $x \in \mathbb{R}^m$ and objective function

$$f(x, \xi) := \begin{pmatrix} \|x - a\| \\ \|x - \xi\| \end{pmatrix}. \tag{37}$$

We consider the simple case of $\xi = (\xi_1, 0, \dots, 0)^T$ with ξ_1 uniformly distributed on $[0, c]$ ($c > 0$ constant). It is easy to see that with $b := (c/2, 0, \dots, 0)^T$, $\mathcal{E}(g(\mathbb{R}^m), \preceq)) = g(\mathrm{conv}\{a, b\})$, i.e. the set of optimal decisions is the line segment joining a and b.

Let us choose $a = (-1, \dots, -1)^T$. This is a convex optimization problem, so it is sufficient to use a smoothing parameter $t = 1$. Denote by k the number of equidistant

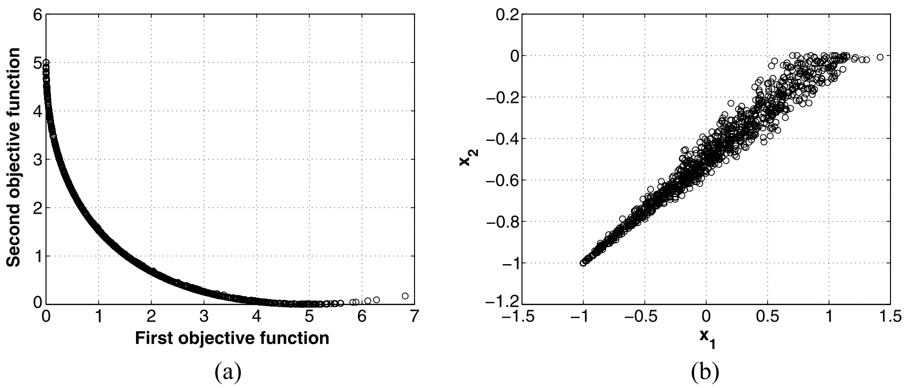
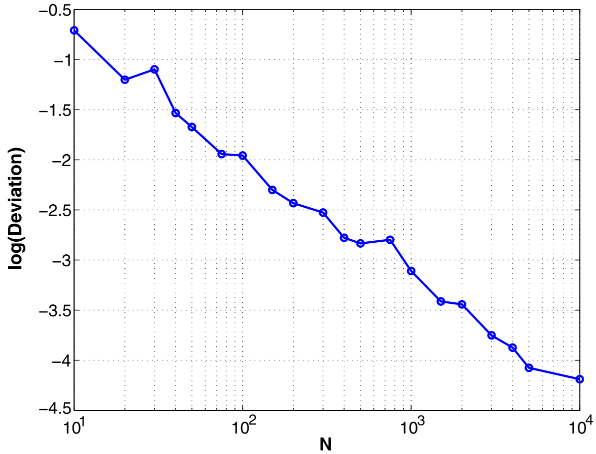
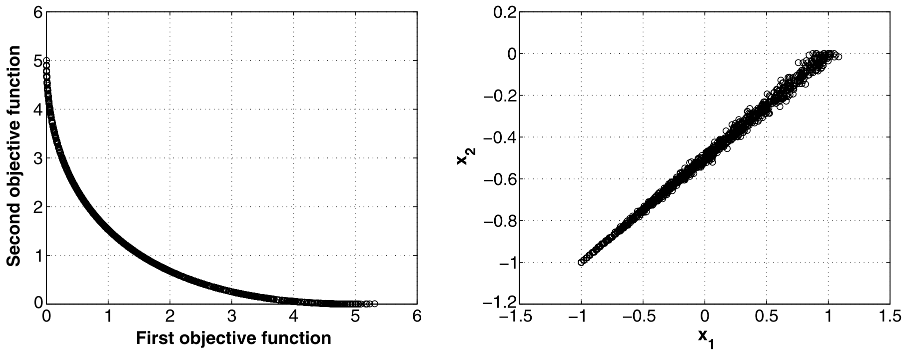


Fig. 1 (**a**) Approximation of the efficient set for problem (37) computed with parameters $k = 100$, $N = 10$, and 10 random starting points per weight vector; (**b**) corresponding preimages

Fig. 2 Sample size N versus deviation between the computed approximation of the set of efficient points and the set of efficient points on a double-logarithmic scale





(a) (b)

Fig. 3 (**a**) Approximation of the efficient set for problem (37) computed with parameters $k = 100$, $N = 100$, and 10 random starting points per weight vector; (**b**) corresponding preimages

weight vectors used from the set $\mathrm{conv}\{(0,1)^T, (1,0)^T\}$. For $m = 2$ and $c = 2$, Fig. 1 depicts the approximation of the efficient set, as well as the corresponding preimages, by 1000 points calculated with $k = 100$ different weights, 10 random starting points per weight vector, and sample size $N = 10$. As it can be seen, the quality of the approximation is already quite high, except for the extreme ends of the curve. The general shape of the solution set $\mathrm{conv}\{(-1,-1)^T, (1,0)^T\}$ is clearly discernible, as is the influence of the random variable on the first coordinate of $b = (1,0)^T$.

Figure 2 shows the deviation between the computed approximation of the set of efficient points and the set of efficient points between $N = 10$ and $N = 10^4$. As it can be seen, the regime of exponential convergence holds over the depicted values of N; see Theorem 5.2.

Let us now increase the sample size to $N = 100$, while keeping all other computational parameters fixed. Figures 3 visualizes the corresponding results. As it can be seen, the quality of the approximation of the optimal decision set has sharply improved, although effects of the random variable are still visible. In contrast to this, the quality of the approximation of the set of efficient points has only slightly improved, doubtless due to the fact that the quality was already quite high for the case $N = 10$. Further experiments with $m \gg 2$ have shown that this effect appears to be relatively independent of the dimension of the underlying decision space.

6.2 A Multiobjective Stochastic Energy Generation Problem

Some methods for multiobjective optimization only work for two criteria [41–43]. Here, we show that our proposed method is also applicable to problems with more than two criteria by considering a game-theoretic problem modified from [44, Example 6.3]: in an electricity market, several electricity generators compete with each other in dispatching electricity. For a total production level $Q \geq 0$, the inverse demand function (which is a price) of generator i is given by $p_i(Q, \xi_i)$, where ξ_i is a random variable. Denote by q_i the electricity production of generator i and let $c_i(q_i)$ be the cost for generator i to produce this amount. Then, every generator wants to maximize his own profit. In [44], this problem has been treated as a noncollaborative stochastic Nash game. Here, however, we treat the problem as a stochastic collaborative game, i.e. we consider the multiobjective stochastic problem with objective

$$f(q,\xi) := \begin{pmatrix} \vdots \\ c_i(q_i) - q_i p_i(\sum_j q_j, \xi_i) \\ \vdots \end{pmatrix} \tag{38}$$

for $q \in \mathbb{R}^m_+$. The connections between collaborative games and MOPs have been well investigated; see, e.g. [28, Chap. 10]. Following [44], we use, for m generators, a random variable $\xi \in \mathbb{R}^m$ where each ξ_i follows a truncated normal distribution with support $[-5, 5]$, mean value 0, and standard deviation 1 ($i = 1, \ldots, m$). With this, we choose $p_i(Q, \xi) = a_i + \xi_i - b_i Q$ and $c_i(q_i) = \alpha_i q_i^2 + \beta_i q_i$, where $a_i, b_i, \alpha_i, \beta_i$ are given constants ($i = 1, \ldots, m$). Note that price functions for different generators may be different. This reflects the fact that, in practice, consumers are willing to pay different prices depending on the nature of how power is generated. We consider the case of $m = 3$ generators and constants $a_i = 25.0, 26.5, 24.0$, $b_i = 1.2, 1.3, 1.5$, $\alpha_i = 5.0, 2.0, 3.0$, and $\beta_i = 1.0, 1.5, 1.2$.

With $t = 0.1$, a sample size of $N = 50$ and 500 randomly generated weight vectors $\omega \in \mathbb{R}^3_{++}$ with $\omega_1 + \omega_2 + \omega_3 = 1$, the corresponding approximation to the set of efficient points is depicted in Fig. 4, while the corresponding preimages are shown in Fig. 5.

Although the objective functions of the problem are not convex, it appears that the set of efficient points corresponds to a subset of a surface of a convex set. As such, Corollary 3.1 ensures that good approximations of the set of efficient points can be constructed for arbitrary t.

6.3 A Nonconvex Academic Testcase

Finally, we consider a nonconvex problem in order to show the behaviour of the approximation scheme proposed on such a seemingly more difficult test case. Following [45], we define a stochastic bi-objective problem with two variables $0 \leq x_1$,

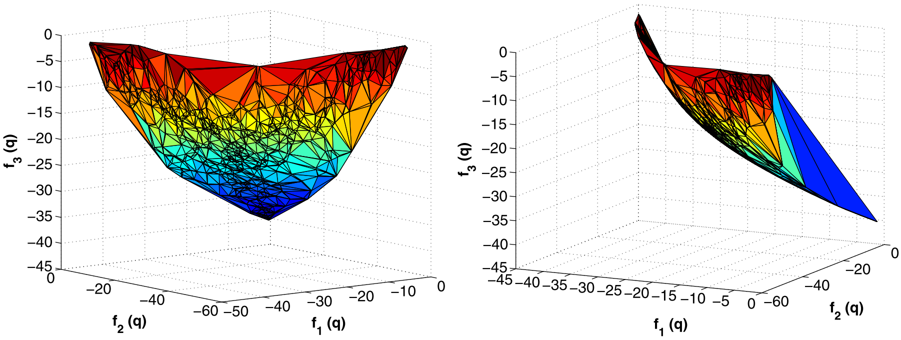


Fig. 4 Approximation of the set of efficient points of problem (38), as seen from two different viewpoints. The triangles of the surface shown correspond to the Delaunay triangulation of the first two coordinates of the points computed

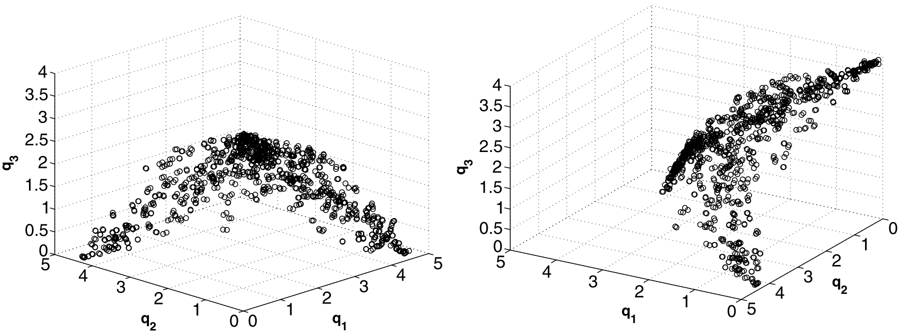


Fig. 5 Preimages of the approximated efficient points of problem (38) in the decision space, as seen from two different viewpoints

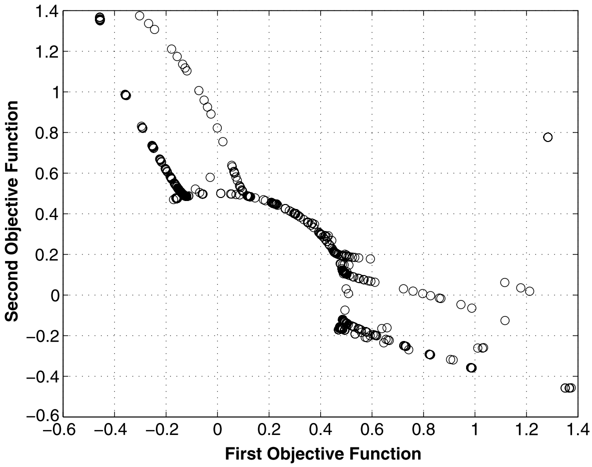
$x_2 \leq 5$, a normally distributed random variable ξ and parametrized objective functions

$$f(x_1, x_2) := \begin{pmatrix} \cos(\frac{2\pi}{360}((a_1 + 10\xi)\sin(2\pi x_1) + h(x_2)))g(x_1) \\ \sin(\frac{2\pi}{360}((a_1 + 10(1-\xi))\sin(2\pi x_1) + h(x_2)))g(x_1) \end{pmatrix}, \tag{39}$$

where $g(x) := 1 + d\cos(2\pi x)$ and $h(x) := a_c + a_2 \sin(2\pi x_2)$. Parameters where chosen as follows: $a_c = 45$, $a_1 = 35$, $a_2 = 20$, and $d = 0.5$.

For this problem, we set $t = 0.1$ for the smoothing parameter, used a sample size of $N = 10$, and used weights $\omega_1 = k/101$, $\omega_2 = 1 - \omega_1$ for $k = 1, \ldots, 100$. The resulting approximation to the set of efficient points is depicted in Fig. 6. Note that, with parameter settings as above, $\mathbb{E}[f(x)]$ is exactly the objective function described by (7.1)–(7.3) in [45]. We can therefore compare Fig. 6 with Fig. 7.1 of [45, p. 110] (also displayed on the title page of [45]) and observe that central parts of the set of efficient points are well approximated, even in the nonconvex region of this set. Note that two branches of locally efficient points are also approximated, which is an artifact of the local optimization solver in use.

Fig. 6 Approximation of the set of efficient points of problem (39)



7 Concluding Remarks

In this paper, we have proposed a smoothing sample average approximation (SAA) method for solving a class of stochastic multiobjective optimization problems where the underlying objective functions are expected values of some random functions. Under some moderate conditions, we have derived the exponential rate of convergence of the SAA problem to its true counterpart in terms of efficient solutions and linear approximation rate in terms of a smoothing parameter. Our preliminary numerical tests show that the method works well. The SAA approach is highly applicable to real life problems where one cannot obtain a closed form of the expected value of a random functions or numerically too expensive to calculate the expected values.

It is possible to extend this work in a number of ways. One is to consider a stochastic multiobjective optimization problem where the objective functions may not appear in a unified form (of the expected of a random function). For instance, in finance, one objective could be the expected profit whereas another could be the variance which represents risks. In such a case, the variance cannot be presented as the expected value of a random function as in (1) and hence our technical results cannot be applied directly to this kind of problem. However, through simple analysis, one can easily find a *nonlinear* form of sample average approximation to the variance and subsequently derive the exponential rate of convergence by following our discussions in Sects. 3–5. We leave this to interested readers.

Another interesting class of stochastic multiobjective optimization problems are those in which the objective functions take the form of the probability of a stochastic constraint. Our conjecture is that we may represent the probability of a stochastic constraint in the form of an expected value of an indicator function and approximate the latter through sample averaging. It is unclear whether it is possible to derive exponential rate of convergence as the functions under the expectation are no longer continuous. It would be interesting to analyse convergence of this type of sample average approximation; this is the subject of future research.

References

1. Goicoechea, A., Hansen, D.R., Duckstein, L.: Multiobjective Decision Analysis with Engineering and Business Applications. Wiley, New York (1982)
2. Stancu-Minasian, I.M.: Stochastic Programming with Multi-Objective Functions. Reidel, Dordrecht (1984)
3. Slowinski, R., Teghem, J. (eds.): Stochastic Versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty. Kluwer Academic, Dordrecht (1990)
4. Caballero, R., Cerda, E., Munoz, M.M., Rey, L.: Efficient solution concepts and their relations in stochastic multiobjective programming. J. Optim. Theory Appl. **110**, 53–74 (2001)
5. Ye, J.J., Zhu, Q.J.: Multiobjective optimization problems with variational inequality constraints. Math. Program. **96**(1), 139–160 (2003)
6. Mordukhovich, B.S.: Multiobjective optimization problems with equilibrium constraints. Math. Program. **117**, 331–354 (2009)
7. White, D.J.: Optimality and Efficiency. Wiley, Chichester (1982)
8. Geoffrion, A.M.: Proper efficiency and the theory of vector maximization. J. Optim. Theory Appl. **22**, 618–630 (1968)
9. Fliege, J.: Gap-free computation of Pareto-points by quadratic scalarizations. Math. Methods Oper. Res. **59**, 69–89 (2004)
10. Fliege, J.: An efficient interior-point method for convex multicriteria optimization problems. Math. Oper. Res. **31**(4), 825–845 (2006)
11. Ben Abdelaziz, F.: L'efficacité en programmation multi-objectifs stochastique. Ph.D. Thesis, Université de Laval, Quebec (1992)
12. Caballero, R., Cerda, E., Munoz, M.M., Rey, L.: Stochastic approach versus multiobjective approach for obtaining efficient solutions in stochastic multiobjective programming problems. Eur. J. Oper. Res. **158**, 633–648 (2004)
13. King, A.J., Wets, R.J.-B.: Epi-consistency of convex stochastic programs. Stoch. Stoch. Rep. **34**, 83–92 (1991)
14. Plambeck, E.L., Fu, B.R., Robinson, S.M., Suri, R.: Sample-path optimization of convex stochastic performances functions. Math. Program. **75**, 137–176 (1996)
15. Robinson, S.M.: Analysis of sample-path optimization. Math. Oper. Res. **21**, 513–528 (1996)
16. Shapiro, A.: Monte Carlo sampling methods. In: Rusczyński, A., Shapiro, A. (eds.) Stochastic Programming. Handbooks in OR & MS, vol. 10. North-Holland, Amsterdam (2003)
17. Xu, H., Meng, F.: Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints. Math. Oper. Res. **32**, 648–668 (2007)
18. De Miguel, V., Xu, H.: A stochastic multiple leader Stackelberg model: analysis, computation, and application. Oper. Res. **57**, 1220–1235 (2009)
19. Wang, W., Ahmed, S.: Sample average approximation of expected value constrained stochastic programs. Oper. Res. Lett. **36**, 515–519 (2008)
20. Peng, J.: A smoothing function and its applications. In: Fukushima, M., Qi, L. (eds.) Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, pp. 293–316. Kluwer Academic, Dordrecht (1998)
21. Hess, C.: Set-valued integration and set-valued probability theory: an overview. In: Handbook of Measure Theory, vols. I, II, pp. 617–673. North-Holland, Amsterdam (2002)
22. Fliege, J., Vicente, L.N.: A multicriteria approach to bilevel optimization. J. Optim. Theory Appl. **131**, 209–225 (2006)
23. Fishburn, P.C.: Mathematics of Decision Theory. Methods and Models in the Social Sciences. Mouton, The Hague (1972)
24. Vogel, W.: Vektoroptimierung in Produkträumen. Systems in Economics, vol. 35. Hain, Meisenheim am Glan (1977)
25. Sawaragi, Y., Nakayama, H., Tanino, T.: Theory of Multiobjective Optimization. Mathematics in Science and Engineering, vol. 176. Academic Press, San Diego (1985)
26. Durier, R.: Weighting factor results in vector optimization. J. Optim. Theory Appl. **58**, 411–430 (1988)
27. Arrow, K.J., Barankin, E.W., Blackwell, D.: Admissible points of convex sets. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games, pp. 481–492. Princeton University Press, Princeton (1953)
28. Jahn, J.: Vector Optimization: Theory, Applications, and Extensions. Springer, Berlin (2004)

29. Göpfert, A., Nehse, R.: Vektoroptimierung—Theorie, Verfahren und Anwendungen. Mathematisch-Naturwissenschaftliche Bibliothek, vol. 74. Teubner, Leipzig (1990)
30. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
31. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer Series in Operations Research. Springer, New York (2000)
32. Yu, P.L.: Cone convexity, cone extreme points, and nondominated solutions in decision problems with multiple objectives. J. Optim. Theory Appl. **14**, 319–377 (1974)
33. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Optimization **57**, 395–418 (2008)
34. Aubin, J.-P., Frankowska, H.: Set-Valued Analysis. Birkhäuser, Boston (1990)
35. Lemaréchal, C.: Bundle methods in nonsmooth optimization. In: Lemaréchal, C., Mifflin, R. (eds.) Nonsmooth Optimization, pp. 79–102. Pergamon Press, Oxford (1978)
36. Rusczyński, A.: A linearization method for nonsmooth stochastic programming problems. Math. Oper. Res. **12**, 32–49 (1987)
37. Dai, L., Chen, C.H., Birge, J.R.: Convergence properties of two stage stochastic programming. J. Optim. Theory Appl. **106**, 489–509 (2000)
38. Shapiro, A.: Asymptotic analysis of stochastic programs. Ann. Oper. Res. **30**, 169–186 (1991)
39. Shapiro, A.: Quantitative stability in stochastic programming. Math. Program. **67**, 99–108 (1994)
40. Römisch, W., Schultz, R.: Lipschitz stability for stochastic programs with complete recourse. SIAM J. Optim. **6**, 531–547 (1996)
41. Das, I., Dennis, J.E.: Normal-boundary intersection: a new method for generating Pareto optimal points in nonlinear multicriteria optimization problems. SIAM J. Optim. **8**, 631–657 (1998)
42. Li, Y., Fadel, G.M., Wiecek, M.M.: Approximating Pareto curves using the hyper-ellipse. In: Proceedings of the 7th AIAA/USAF Conference (1998)
43. Ralphs, T.K., Saltzman, M.J., Wiecek, M.M.: An improved algorithm for solving bi-objective integer programs. Ann. Oper. Res. **147**, 43–70 (2006)
44. Xu, H., Zhang, D.: Stochastic Nash equilibrium problems: sample average approximation and applications. Available at http://www.optimization-online.org (2009)
45. Hillermeier, C.: Nonlinear Multiobjective Optimization: A Generalized Homotopy Approach. Birkhäuser, Basel (2001)