Scenario Approximation of Robust and Chance-Constrained Programs

Raffaello Seri & Christine Choirat

Journal of Optimization Theory and Applications

ISSN 0022-3239
Volume 158
Number 2

J Optim Theory Appl (2013) 158:590-614
DOI 10.1007/s10957-012-0230-3

Springer

J Optim Theory Appl (2013) 158:590–614
DOI 10.1007/s10957-012-0230-3

Scenario Approximation of Robust and Chance-Constrained Programs

Raffaello Seri · Christine Choirat

Received: 8 November 2012 / Accepted: 15 November 2012 / Published online: 9 January 2013

Abstract We consider scenario approximation of problems given by the optimization of a function over a constraint that is too difficult to be handled but can be efficiently approximated by a finite collection of constraints corresponding to alternative scenarios. The covered programs include min-max games, and semi-infinite, robust and chance-constrained programming problems. We prove convergence of the solutions of the approximated programs to the given ones, using mainly epigraphical convergence, a kind of variational convergence that has demonstrated to be a valuable tool in optimization problems.

Keywords Mathematical programming · Epigraphical convergence · Scenario approximation · Sampling

1 Introduction

The paper aims at providing general conditions under which some programming problems subject to constraints can be approximated by *scenario programs*. The covered programming problems include min-max games, and semi-infinite, robust and chance-constrained programs. All of these problems can be written as the optimization of a function under a constraint that is too difficult to be handled. However, it can be efficiently approximated by a finite collection of constraints corresponding

Communicated by Masao Fukushima.

R. Seri (✉)
Dipartimento di Economia, Università degli Studi dell'Insubria, Varese, Italy
e-mail: raffaello.seri@uninsubria.it

C. Choirat
Department of Economics, School of Economics and Business Management, Universidad de Navarra, Pamplona, Spain
e-mail: cchoirat@unav.es

to alternative scenarios, which can be either the result of a deterministic approximation method or the realization of a sequence of random variables. Several recent papers deal either with the computational complexity of these algorithms or with the precision of these methods (see, e.g., [1–6]). We address another problem: finding conditions ensuring convergence of the solutions of the scenario programs to the solution of the original ones, when the number of scenarios increases to infinity. This is indeed a minimal requirement to be checked before evaluating the complexity and the accuracy of the method.

In the case of semi-infinite programming, these algorithms are known as *discretization methods* or *outer approximation methods* (see [7] for an introduction to the topic), and their convergence properties have been investigated in several papers (see, e.g., [7–10]). However, these convergence results either are stated in an abstract setting [7, 9] that requires a further verification of the hypotheses, or are focused [8, 10] and can be applied only to specialized (and efficient) algorithms. The results provided in this paper stand in-between the two approaches, and benefit from the use of a kind of convergence (epigraphical convergence; see below) more general than customarily used. At last, the framework we propose is not limited to semi-infinite programming, and some insight in the conditions ensuring convergence can be gained from the joint analysis of these situations.

2 Preliminaries

It is important to stress the similarities and the differences of scenario programs with respect to the *sample average approximation scheme* used in Stochastic Programming (see [11] and the references therein). In Stochastic Programming, the objective is the maximization of an integral functional over a fixed set, and the integral functional is approximated through an empirical mean functional while the set of constraints is kept fixed. On the other hand, in the present case, the objective function is known, while the set has to be approximated (see [4, Sect. 2.1.5]). This allows the use, in Stochastic Programming, of a large machinery developed in Statistics to deal with the class of so-called M-estimators (see, e.g., [12]) and leaves Robust Programming with the open problem of developing methods for the proof of asymptotic properties.

As a result, our theorems show that the solutions of the scenario programs approach the solutions of the original ones under conditions that appear to be new in the literature and more technical than in the case of Stochastic Programming. In passing, we stress why the customary definition of a class of stochastic semi-infinite programs is unsuitable and we reformulate it (see Eq. (3)).

In order to prove convergence of the solutions of the approximated programs to the original ones, we use mainly epigraphical convergence, a kind of variational convergence that has demonstrated to be a valuable tool in optimization problems. As stressed above, our main objective is to show convergence of optimal solutions, and not to derive feasibility of programming algorithms: therefore, our results will not involve, as is customary in programming, convexity of sets and functions, but, e.g., closedness of sets and semi-continuity of functions. These requirements will often be a fortiori true under the standard hypotheses, but they allow one to deal with much more general situations.

The contents of the paper can be summarized as follows. In Sect. 3, we describe the programming problems considered in the rest of the paper, namely semi-infinite, robust and chance-constrained programs. The relative approximation schemes are dealt with in Sect. 4, while Sect. 5 contains the statements of the main results concerning consistency of the previous scenario approximations. A numerical example is studied in Sect. 6. Lastly, Sect. 7 concludes, Appendix A contains the proofs of the results and Appendix B collects the results on epi-convergence, stationary and ergodic sequences and random sets that are used in the paper.

We end with some notational remarks. Let $\mathbb{N}$ and $\mathbb{R}$ denote, respectively, the set of natural and real numbers. In general, small letters (x) stand for variables, capital letters (X) for random variables, bold letters ($\mathbf{X}$) for abstract spaces, and calligraphic letters ($\mathcal{X}$) for σ-algebras (or more general collections of sets, if otherwise specified). We suppose that the optimization variable x takes its values in the feasible set $\mathbf{X} \subseteq \mathbb{R}^p$ and represents the variables on which the individual has a direct control. On the other hand, $y \in \mathbf{Y}$ is a variable that cannot be chosen by the individual and that can, in some cases, be the realization of a random variable (then written as Y). A sequence of elements is denoted as $(x^{(i)})_{i=1,\dots,n} = (x^{(i)})_i$ where i is the index of the sequence; a point set is denoted as $\{x^{(i)}\}$ (see [13, p. 14] for an explanation of the difference in Numerical Analysis). The difference is that a (finite) sequence of length n is obtained selecting the elements from 1 to n of an infinite sequence and is therefore expanded through the addition of new points; on the other hand, this does not hold for a point set. The results in the following are stated for sequences since this is the most natural framework, especially when dealing with stochastic methods. They can be adapted to cover the case of points sets, as the numerical example in Sect. 6 will show.

3 Description of the Problem

We start from the following *min-max game*:

$$\min_{x \in \mathbf{X}} \max_{y \in \mathbf{Y}} g(x, y), \tag{1}$$

for a function $g : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}$. The rationale behind a min-max game is that the decision maker wants to choose a value of x that minimizes the loss g arising from his choice in the worst possible scenario, i.e. when his opponent chooses the value y in order to maximize the loss. It is simple to remark that this can be written as

$$\min_{(\gamma,x) \in \mathbb{R} \times \mathbf{X}} \gamma, \quad \text{subject to} \quad g(x, y) - \gamma \leq 0, \quad \forall y \in \mathbf{Y}.$$

This is an example of a more general *semi-infinite program* (see, e.g., [14, 15]):

$$\min_{x \in \mathbf{X}} h(x), \quad \text{subject to} \quad f(x, y) \leq 0, \quad \forall y \in \mathbf{Y}. \tag{2}$$

The link between semi-infinite programs and min-max problems is investigated in [16]. Apart from the case of min-max games, Eq. (2) often emerges from a *robust feasibility program*, when the aim is to find a value of $x \in \mathbf{X}$ (if it exists) such that $f(x, y) \leq 0$ for any $y \in \mathbf{Y}$.

A stochastic counterpart of a min-max game and of a semi-infinite program is obtained when y is the realization of a random variable defined on the probability space $(\mathbf{Y}, \mathcal{Y}, \mathbb{P})$. This problem is customarily written in the literature as

$$\min_{x \in \mathbf{X}} h(x), \quad \text{subject to} \quad f(x, Y) \leq 0$$

where Y is a random variable with probability $\mathbb{P}$. Even if intuitively appealing, this formulation is nevertheless not correct. Without entering the mathematical details (see [17] for a full derivation), the main problem of this formulation is that it requires the constraint to hold for any value assumed by Y: however, random variables are only defined up to the values taken on null sets, so that it is not possible to guarantee that the constraint is in fact respected on these null sets. To provide a concise and correct statement of the problem, we introduce the *continuous intersection*[1] (see, e.g., [18, p. IV-34]) of a random set Γ. Let $\Gamma : y \mapsto \Gamma(y)$ be a random set (see Appendix B), parameterized by the random element y. Then the continuous intersection of Γ is

$$\operatorname{Int}(\Gamma) := \bigcup_{N \in \mathcal{N}} \bigcap_{y \in \operatorname{supp}(\mathbb{P}) \setminus N} \Gamma(y),$$

where $\mathcal{N}$ denotes the set of all null sets of $(\mathbf{Y}, \mathcal{Y}, \mathbb{P})$, and $\operatorname{supp}(\mathbb{P})$ is the support of $\mathbb{P}$, that is the smallest closed subset of $\mathbf{Y}$ with full $\mathbb{P}$-measure. Therefore, we define a *robust program* as

$$\min_{x \in \mathbf{X}} h(x), \quad \text{for } x \in \operatorname{Int}\bigl(\bigl\{x \in \mathbf{X} : f(x, Y) \leq 0\bigr\}\bigr), \tag{3}$$

where $f : \mathbf{X} \times \mathbf{Y} \to \mathbb{R}$. Indeed, (3) can be equivalently written as the minimization of $h(x)$ for $x \in \mathbf{X}$, such that there exists $N_x \in \mathcal{Y}$ for which $\mathbb{P}(N_x) = 0$ (i.e. $N_x \in \mathcal{N}$) and $f(x, y) \leq 0$ for any $y \in \mathbf{Y} \setminus N_x$.

Robust programs arise from Eq. (1), when the choice of the opponent (that is, $y \in \mathbf{Y}$) can be considered a random variable, as is the case when the opponent is "Nature". The difference between the two programs (2) and (3) is largely arbitrary since it is often possible to define a probability measure on $\mathbf{Y}$ such that the solution of (2) almost surely coincides with the solution of (3), as our results will show. However, they differ since the most natural approximation strategy is different in the two cases. While both (2) and (3) can be approximated through algorithms based on deterministic sequences (see Theorem 5.1), random sampling (see Corollary 5.2) is more suitable for (3).

Another model that can be considered as a stochastic version of (2), is the following kind of program, introduced in [19–21] (see also [22], for the relevance of distributional assumption in this framework), and often called a *probabilistic* or *chance-constrained program*, or *program with probabilistic* or *chance-constraints*:

$$\min_{x \in \mathbf{X}} h(x), \quad \text{subject to} \quad \mathbb{P}\bigl\{(x, Y) \in A\bigr\} \geq \alpha, \tag{4}$$

[1] This object is also called *essential intersection*, but this name is quite misleading since it induces some confusion with the $\mathbb{P}$-*essential intersection*.

where A is a subset of $\mathbf{X}\times\mathbf{Y}$. It is always possible to write the set A as $A := \{(x,y) \in \mathbf{X}\times\mathbf{Y} : f(x,y) \le 0\}$, thus offering a striking similarity with program (3).[2]

Remark 3.1 The programs previously exposed admit several extensions. In the following, we show how these generalizations[3] can be dealt with in our framework.

(i) Remark that in (2) we could consider without loss of generality a linear cost function $h(x) = c^{\mathrm{T}}x$, since (2) can be written in the form

$$\min_{(\gamma,x)\in\mathbb{R}\times\mathbf{X}} \gamma, \quad \text{subject to} \quad h(x)-\gamma \le 0 \text{ and } f(x,y) \le 0, \quad \forall y \in \mathbf{Y}. \tag{5}$$

Therefore, in the following, we will use this simpler form. In Remark 5.1(iii), the alternative case of a generic function $h(\cdot)$ is also treated; this shows that the assumption of linearity has an interest in the statement of the theorems since it requires a more careful analysis of level-boundedness.

(ii) Consider the case in which the constraints of (3) are given by the inequalities $f_i(x, Y_i) \le 0$ for $i = 1, \dots, m$, where any Y_i takes its values in the probability space $(\mathbf{Y}_i, \mathcal{Y}_i, \mathbb{P}_i)$. We define a new random variable $Y = (Y_1, \dots, Y_m)$ taking values in the space $\mathbf{Y} = \prod_{i=1}^m \mathbf{Y}_i$ endowed with the product σ-algebra ($\mathcal{Y} = \bigotimes_{i=1}^m \mathcal{Y}_i$) and an adequate probability measure $\mathbb{P}$ (this allows one to consider dependence among the components of Y, but if the random variables are independent we take $\mathbb{P} = \bigotimes_{i=1}^m \mathbb{P}_i$). Define moreover a projection operator as $e_i : \mathbf{Y} \to \mathbf{Y}_i$ such that $e_i(Y) = Y_i$. Then $f_i(x, Y_i) = f_i(x, e_i(Y))$ and this program can be written as in (3) defining $f(x,Y) = \min_{i=1,\dots,m} f_i(x, e_i(Y))$.

(iii) Still (see [1, 23, 24]) considers *generalized semi-infinite programs* or *semi-infinite programs with variable index sets*, defined by

$$\min_{x\in\mathbf{X}} h(x), \quad \text{subject to} \quad x \in \bigl\{x \in \mathbf{X} : f(x,y) \le 0, \forall y \in \mathbf{Y}(x)\bigr\}, \tag{6}$$

where the index sets $\mathbf{Y}(x) \subseteq \mathbb{R}^\ell$ are allowed to depend on $x \in \mathbf{X}$. Also in this case, the function f can be substituted with a linear form as in (5); more interesting is the fact that also the variable index sets can be removed. Define the new functions $f^\star(x,y) = f(x,y) + \chi(y, \mathbf{Y}(x))$, where χ is the *indicator function* of convex analysis.[4] Then problem (6) becomes a classical semi-infinite programming problem:

$$\min_{x\in\mathbf{X}} h(x), \quad \text{subject to} \quad f^\star(x,y) \le 0, \quad \forall y \in \mathbf{Y}.$$

Clearly, $f^\star$ inherits the properties of f and $\mathbf{Y}(\cdot)$.

[2] It is enough to take $f(x,y) = 1 - 2\cdot \mathbf{1}\{(x,y) \in A\}$, where $\mathbf{1}\{z \in B\}$ is the *indicator* or *characteristic function*, that is the function taking the value 1 if $z \in B$ and 0 otherwise.

[3] Some of them are discussed in [3, pp. 99–100].

[4] χ is defined as:

$$\chi(x,C) = \begin{cases} 0, & \text{if } x \in C \\ +\infty, & \text{if } x \notin C \end{cases}, \qquad x \in \mathbf{X}.$$

(iv) In [25], the authors consider the case in which (3) is substituted by

$$\min_{x\in\mathbf{X}} \int_{\mathbf{Y}} g(x,y)\mathbb{P}(\mathrm{d}y) \quad \text{for } x \in \operatorname{Int}\bigl(\bigl\{x \in \mathbf{X} : f(x,Y) \le 0\bigr\}\bigr).$$

Note that we are using the corrected version of the constraint (see Eq. (3)). In this case, a double approximation strategy can be used, using some points in $\mathbf{Y}$ to approximate both the function $\int_{\mathbf{Y}} g(x,y)\mathbb{P}(\mathrm{d}y)$ and the set $\operatorname{Int}(\{x \in \mathbf{X} : f(y,x) \le 0\})$ (see, e.g., [26] for the theory, and [27] for an application). It is unclear whether it is better to use the same points in the two tasks or to base the two approximations on two different sequences. Nevertheless, the convergence of the objective function can be obtained along the lines, e.g., of [28, 29], and the convergence of the constraints follows from the results discussed in the remainder of the paper. Once the epi-convergence of the objective function and of the constraints are obtained, the results on epi-convergence of sums in Chap. 6 of [30] can be used.

(v) In (4), it is often possible to introduce several bounds on the probabilities (see [31]). This kind of programs too can be dealt with adapting the following results.

4 Approximation Schemes

Consider the semi-infinite program (2) (as modified according to Remark 3.1):

$$\min_{x\in\mathbf{X}} c^{\mathrm{T}}x, \quad \text{subject to} \quad f(x,y) \le 0, \quad \forall y \in \mathbf{Y}. \tag{7}$$

It is sensible to approximate it through the discretization approximation:

$$\min_{x\in\mathbf{X}} c^{\mathrm{T}}x, \quad \text{subject to} \quad f\bigl(x,y^{(i)}\bigr) \le 0, \quad i = 1,\dots,n,$$

where the values $(y^{(i)})_{i=1,\dots,n}$ are chosen in a deterministic way (see, e.g., [7, 32]).

A similar approximation was also proposed in [2–4] to approximate the robust optimization problem of Eq. (3):

$$\min_{x\in\mathbf{X}} c^{\mathrm{T}}x, \quad \text{for } x \in \operatorname{Int}\bigl(\bigl\{x \in \mathbf{X} : f(x,Y) \le 0\bigr\}\bigr) \tag{8}$$

through the sample approximation

$$\min_{x\in\mathbf{X}} c^{\mathrm{T}}x, \quad \text{subject to} \quad f\bigl(x,y^{(i)}\bigr) \le 0, \quad i = 1,\dots,n,$$

where the values $(y^{(i)})_{i=1,\dots,n}$ are now identically and independently sampled according to the probability $\mathbb{P}$. The authors call this the *scenario program*.[5]

Summing up, we consider the following approximate program:

$$\min_{x\in\mathbf{X}} c^{\mathrm{T}}x, \quad \text{subject to} \quad f\bigl(x,y^{(i)}\bigr) \le 0, \quad i = 1,\dots,n, \tag{9}$$

[5] This is similar to the program considered in [25].

where the points $(y^{(i)})_{i=1,\dots,n}$ may be a sequence of either deterministic or stochastic points.[6] The aim of this paper is to show that, under certain conditions on the sample $(y^{(i)})_{i=1,\dots,n}$ and on the program, the solution to (9) converges to the solution of (7) or (8) as long as $n \to \infty$. Clearly, convergence will be replaced by almost sure convergence in the stochastic case.

As concerns the chance-constrained program (4), the approximation works as follows. If we denote by $\mathbf{1}\{B\}$ the *indicator* or *characteristic function* of the set B (i.e. the function taking the value 1 for $x \in B$ and 0 for $x \notin B$), the probability in (4) can be approximated as

$$\mathbb{P}\big\{(x,Y) \in A\big\} = \mathbb{E}\mathbf{1}\big\{(x,Y) \in A\big\} \cong \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\big\{\big(x, y^{(i)}\big) \in A\big\}$$

for a sequence of points $(y^{(i)})_{i=1,\dots,n}$. Moreover, we suppose that also the value α appearing in (4) is function of n (the reason for this choice will be clear in the following). Therefore we are led to consider the new program (or better, the class of programs indexed by the sequence $(\alpha_n)_n$):

$$\min_{x \in \mathbf{X}} h(x), \quad \text{subject to} \quad \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\big\{\big(x, y^{(i)}\big) \in A\big\} \geq \alpha_n. \tag{10}$$

Now we introduce some definitions that will be useful in the following (see, e.g., [33]).

Definition 4.1 For a function $f : \mathbb{R}^k \to \mathbb{R}$, we define the *effective domain* as $\operatorname{dom} f := \{x \in \mathbb{R}^k : f(x) < \infty\}$, and the level sets as

$$\operatorname{lev}_{\leq \alpha} f := \big\{x \in \mathbb{R}^k : f(x) \leq \alpha\big\},$$

for any $\alpha \in \mathbb{R}$; it is possible to define similarly the sets $\operatorname{lev}_{<\alpha} f$, $\operatorname{lev}_{=\alpha} f$, $\operatorname{lev}_{>\alpha} f$ and $\operatorname{lev}_{\geq \alpha} f$. A function f is said to be *level-bounded* iff, for every $\alpha \in \mathbb{R}$, the set $\operatorname{lev}_{\leq \alpha} f$ is bounded. A function f is *proper* iff $f(x) < \infty$ for at least one $x \in \mathbb{R}^k$, and $f(x) > -\infty$ for all $x \in \mathbb{R}^k$; it is *lower semi-continuous* (lsc) at $\overline{x}$ iff $\liminf_{x \to \overline{x}} f(x) \geq f(\overline{x})$; it is *lower semi-continuous* iff it is lsc at any $\overline{x} \in \mathbb{R}^k$ or, equivalently, iff its level sets $\operatorname{lev}_{\leq \alpha} f$ are all closed in $\mathbb{R}^k$. The sequence $(f_n)_{n \in \mathbb{N}}$ is *eventually level-bounded* iff there is a level-bounded function g such that eventually (i.e. for any n large enough) $f_n \geq g$, or iff the sequence of sets $\operatorname{dom} f_n$ is eventually bounded.

[6] It would be clearly possible to explicitly consider two different approximating programs, one for the deterministic and one for the stochastic case. However, the separation between the two programs is quite artificial. As an example, also for (7) it is possible to define a fictitious probability measure on $\mathbf{Y}$ and to draw random points according to it. Provided the density of the measure is strictly positive, the behavior of the solutions is described by Corollary 5.2.

5 Main Results

5.1 Semi-infinite Programming

The following theorem gives conditions under which the solution of (9) converges to the solution of a limit program that can coincide with the solution of (7). Clearly, nothing guarantees that the approximate problem (9) has indeed a solution for finite n, but the theorem yields conditions under which such a solution asymptotically exists.

Theorem 5.1 *Let $f : \mathbf{X} \times \mathbf{Y} \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous in x for any $y \in \mathbf{Y}$, and suppose that the set $\bigcap_{y \in \mathbf{Y}} \{x \in \mathbf{X} : f(x, y) \le 0\}$ is nonempty. Let $\overline{x}_n$ denote a solution to the program (9); then the cluster points of the sequence $(\overline{x}_n)_n$ belong to the solution set of*

$$\min_{x \in \mathbf{X}} c^{\mathrm{T}} x, \quad \textit{subject to} \quad f(x, y) \le 0, \quad \forall y \in \mathbf{Y}^{\star}, \tag{11}$$

where $\mathbf{Y}^{\star} = \{y^{(i)}, i \in \mathbb{N}\}$, provided that there exists an index n_0 such that the set $\bigcap_{i=1}^{n} \{x \in \mathbf{X} : f(x, y^{(i)}) \le 0\}$ is compact for any $n \ge n_0$.

If (11) has just one solution $\overline{x}$, then $\lim_{n \to \infty} \overline{x}_n = \overline{x}$.

Remark 5.1 (i) The theorem does not guarantee that the solution of (7) and the limiting solution of (9) coincide. This fact depends on the set $\mathbf{Y}^{\star}$, whose choice, apart from the obvious requirement of denumerability, is entirely left to the researcher. However, Corollary 5.1 shows that it is convenient to devise algorithms such that $\mathbf{Y}^{\star}$ is dense in $\mathbf{Y}$. This requirement is equivalent to the fact that the discretization mesh-size of $(y^{(i)})_{i=1,\ldots,n}$ converges to 0 as the number of points n increases. The rate of convergence of the solutions of (9) to the solution of (7) as a function of the discretization mesh-size is investigated in [1].
(ii) The eventual compactness of $\bigcap_{i=1}^{n} \{x \in \mathbf{X} : f(x, y^{(i)}) \le 0\}$ is only used to prove the existence of the cluster points and can be avoided if the result is restated as follows: the cluster points of the sequence $(\overline{x}_n)_n$, provided they exist, belong to the solution set of (11) where $\mathbf{Y}^{\star} = \{y^{(i)}, i \in \mathbb{N}\}$ (see [17] for a proof).
(iii) It is possible to replace $c^{\mathrm{T}} x$ with any continuous and proper function $h(\cdot)$. In this case, the condition on the existence of an n_0 such that the set

$$\bigcap_{i=1}^{n} \left\{x \in \mathbf{X} : f\left(x, y^{(i)}\right) \le 0\right\}$$

is compact for any $n \ge n_0$ can be avoided if the function $h(\cdot)$ is also level-bounded. What is important is that either $h(\cdot)$ is level-bounded or the space over which the min is taken is eventually bounded.

The following corollary gives a condition under which the solution of (11) coincides with the solution of (7).

Corollary 5.1 *Let* $\mathbf{Y}$ *be a metric space and let* $f:\mathbf{X}\times\mathbf{Y}\to\mathbb{R}\cup\{+\infty\}$ *be lower semi-continuous in* y *for any* $x\in\mathbf{X}$. *Then, if* $\mathbf{Y}^\star$ *is a dense subset of* $\mathbf{Y}$, *the following equality holds*:

$$\bigcap_{y\in\mathbf{Y}^\star}\big\{x\in\mathbf{X}: f(x,y)\le 0\big\}=\bigcap_{y\in\mathbf{Y}}\big\{x\in\mathbf{X}: f(x,y)\le 0\big\},$$

and the solution of (11) *coincides with the solution of* (7).

5.2 Robust Programming

The following corollary shows how Theorem 5.1 can be applied when the points are randomly (but not necessarily independently) drawn according to a probability distribution $\mathbb{P}$.

Corollary 5.2 *Suppose that* $(y^{(i)})_i$ *is the realization of an ergodic and strictly stationary sequence of points* (*see Appendix* B *for definitions*) *defined on the probability space* $(\mathbf{Y},\mathcal{Y},\mathbb{P})$. *Suppose*, *moreover, that*:

(i) $f:\mathbf{X}\times\mathbf{Y}\to\mathbb{R}\cup\{+\infty\}$ *is lower semi-continuous in* x *for* $\mathbb{P}$*-almost surely any* $Y\in\mathbf{Y}$;
(ii) *the set* $\{x\in\mathbf{X}: f(x,Y)\le 0\}$ *is nonempty for* $\mathbb{P}$*-almost surely any* $Y\in\mathbf{Y}$;
(iii) *the random variable* $D(Y):=d(0,\{x\in\mathbf{X}: f(x,Y)\le 0\})$ *is* $\mathbb{P}$*-integrable.*

Let $\overline{x}_n$ *denote the solution to the program* (9); *then*, *under the hypotheses of Theorem* 5.1, *the cluster points of the sequence* $(\overline{x}_n)_n$ *almost surely belong to the solution of*

$$\min_{x\in\mathbf{X}} c^{\mathrm{T}}x,\quad \textit{for } x\in\operatorname{Int}\big(\big\{x\in\mathbf{X}: f(x,y)\le 0\big\}\big). \tag{12}$$

The value $\lim_{n\to\infty}c^{\mathrm{T}}\overline{x}_n$ *is almost surely equal to the value of the solution of* (8). *Moreover, if* (12) *has just one solution* $\overline{x}$, *then* $\lim_{n\to\infty}\overline{x}_n=\overline{x}$ $\mathbb{P}$*-almost surely.*

Remark 5.2 (i) Remark 5.1 (iii) applies also in this context.
(ii) In this corollary, the points $(y^{(i)})_i$ are allowed to be the realization of a stationary ergodic sequence: this can be useful, for example, when the $(y^{(i)})_i$ are obtained through the observation of a real world situation (see, e.g., [34], for the case of chance-constrained programming). The extension to asymptotically mean stationary sequences (see [35]) is also possible along the lines of [29].
(iii) Remark that, if the probability measure $\mathbb{P}$ is discrete, then the equality $\operatorname{Int}(\{x\in\mathbf{X}: f(x,y)\le 0\})=\bigcap_{y\in\mathbf{Y}}\{x\in\mathbf{X}: f(x,y)\le 0\}$ holds true.[7]

[7] This can be simply verified using the result of [18] quoted in the proof of Corollary 5.2. Indeed, if (p_y) are the probability masses, we have

$$\begin{aligned}\chi\big(x,\operatorname{Int}\big\{x\in\mathbf{X}: f(y,x)\le 0\big\}\big)&=\sum_{y\in\mathbf{Y}}\chi\big(x,\big\{x\in\mathbf{X}: f(y,x)\le 0\big\}\big)\cdot p_y\\&=\chi\Big(x,\bigcap_{y\in\mathbf{Y}}\big\{x\in\mathbf{X}: f(y,x)\le 0\big\}\Big).\end{aligned}$$

(iv) Some results concerning the quality of the scenario approximation in the case of independent and identically distributed points are given in [2–4, 6].

5.3 Chance-Constrained Programming

The following theorem contains a result concerning convergence of optimal solutions of approximated chance-constrained programs to the optimal solutions of the original ones. Statements *(a)* and *(b)* essentially correspond to Theorem 4.5 of [36] and use epi-convergence. As concerns statement *(c)*, the first part is inspired by Proposition 2.2 in [37] and uses epi-convergence, while in order to derive the second part, we use uniform convergence. We will need the following definition.

Definition 5.1 A class of sets $\mathcal{A}$ is said to be $\mathbb{P}$-Glivenko–Cantelli (see, e.g., [38]) iff:

$$\sup_{A\in\mathcal{A}} \big|\mathbb{P}_n\{Y\in A\}-\mathbb{P}\{Y\in A\}\big| \underset{n\to\infty}{\longrightarrow} 0 \quad \mathbb{P}\text{-as},$$

where $\mathbb{P}_n$ is an empirical probability based on a sample $(Y^{(i)})_i$ from $\mathbb{P}$, i.e. $\mathbb{P}_n\{Y\in A\}=\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{Y^{(i)}\in A\}$.

Example 5.1 Let $\mathbb{P}$ be any probability measure on $\mathbb{R}^k$. Then the class of sets $\mathcal{A}=\{\prod_{i=1}^k(-\infty,x_i], x=(x_1,\dots,x_k)\in\mathbb{R}^k\}$ is $\mathbb{P}$-Glivenko–Cantelli. More general examples can be found, e.g., in [38, 39].

Theorem 5.2 *Suppose that $(y^{(i)})_i$ is a sequence of independent and identically distributed points defined on the probability space $(\mathbf{Y},\mathcal{Y},\mathbb{P})$, and consider problem (10). Suppose, moreover, that*:

(i) *The set A belongs to $\mathcal{B}(\mathbf{X})\otimes\mathcal{Y}$, where $\mathcal{B}(\mathbf{X})$ is the Borel σ-algebra of $\mathbf{X}$, and is such that, for $\mathbb{P}$-almost any realization y of Y, $A\cap(X\times\{y\})$ is closed.*
(ii) *The function $h(x)$ is continuous, proper and level-bounded.*
(iii) *For $\mathbb{P}$-almost any sequence $(y^{(i)})_i$, there exists an index n_0 (depending on the sequence $(y^{(i)})_i$) such that the set*

$$\left\{x\in\mathbf{X}:\frac{1}{n}\sum_{i=1}^n \mathbf{1}\big\{\big(x,y^{(i)}\big)\in A\big\}\geq\alpha_n\right\}$$

is nonempty and compact for any $n\geq n_0$.
(iv) *The set $\{x\in\mathbf{X}:\mathbb{P}\{(x,Y)\in A\}\geq\alpha_n\}$ is nonempty.*
(v) *$\{A(x),x\in\mathbf{X}\}$, where $A(x):=\{y\in\mathbf{Y}:(x,y)\in A\}$, is a Glivenko–Cantelli collection of sets.*
(vi) *The following inclusion holds*:

$$\big\{x\in\mathbf{X}:\mathbb{P}\big\{(x,Y)\in A\big\}\geq\alpha_n\big\}\subseteq\mathrm{cl}\big[\big\{x\in\mathbf{X}:\mathbb{P}\big\{(x,Y)\in A\big\}>\alpha_n\big\}\big],$$

where $\mathrm{cl}(A)$ is the closure of the set A.

(a) *Let* $\overline{x}_n$ *denote a solution to the program* (10); *then, for any sequence* $(\alpha_n)_n$ *respecting hypotheses* (i)–(ii)–(iv), *the cluster points of the sequence* $(\overline{x}_n)_n$ *are* $\mathbb{P}$*-almost surely feasible* (*but not necessary optimal*) *solutions of* (4). *Moreover, for any sequence* $(\alpha_n)_n$ *respecting hypotheses* (i)–(ii)–(iii)–(iv), *we have that* $\liminf_n h(\overline{x}_n) \geq h(\overline{x})$ $\mathbb{P}$*-almost surely.*
(b) *Moreover, there exists a sequence* $(\alpha_n^\star)_n$ *with* $\alpha_n^\star \uparrow \alpha$ *such that, if this sequence respects hypotheses* (i)–(ii)–(iii)–(iv), *then the cluster points of the sequence* $(\overline{x}_n)_n$ $\mathbb{P}$*-almost surely belong to the solution of* (4). *If* (4) *has just one solution* $\overline{x}$, *then* $\lim_{n\to\infty} \overline{x}_n = \overline{x}$ $\mathbb{P}$*-almost surely.*
(c) *Let* $\overline{x}_n$ *denote a solution to the program* (10) *with* $\alpha_n = \alpha$. *If hypotheses* (i)–(ii)–(iii)–(iv)–(vi) *hold, then* $\lim_{n\to\infty} h(\overline{x}_n) = h(\overline{x})$ $\mathbb{P}$*-almost surely. If hypotheses* (i)–(ii)–(iii)–(v)–(vi) *hold, the cluster points of the sequence* $(\overline{x}_n)_n$ $\mathbb{P}$*-almost surely belong to the solution of* (4); *if* (4) *has just one solution* $\overline{x}$, *then* $\lim_{n\to\infty} \overline{x}_n = \overline{x}$ $\mathbb{P}$*-almost surely.*

Remark 5.3 (i) The Monte Carlo sampling algorithm can be replaced by any other probability approximation algorithm, provided the approximated probability converges epigraphically (for (a), (b) and the first part of (c)) or uniformly (for the second part of (c)) to the original one. In particular, results (a), (b) and the first part of (c) hold under the same hypotheses, if the points $(y^{(i)})_i$ come from a strictly stationary and ergodic sequence.
(ii) Hypothesis (v) requires $\{A(x), x \in \mathbf{X}\}$ to be a Glivenko–Cantelli collection of sets. In the case of a sequence of independent and identically distributed points, this can be assessed showing that the class of sets respects the conditions in [39, p. 119] on the entropy, or that the class is a Vapnik–Chervonenkis one (see [40, pp. 827ff.]). Proposition 2.1 in [37] provides an interesting condition under which hypothesis (v) holds with explicit reference to chance constraints. In [41], some useful references to Glivenko–Cantelli results for collections of sets in the case of dependent sequences are provided. The relation between Vapnik–Chervonenkis dimension and accuracy of approximation is discussed in [3, Remark 1 in Sect. 2.1] and [4, pp. 28–29].
(iii) Hypothesis (vi) can be interpreted as a continuity condition on the probability $\mathbb{P}\{(x, Y) \in A\}$ around the value α. Otherwise stated, it requires that, for any point $\overline{x}$ in $\{x \in \mathbf{X} : \mathbb{P}\{(x, Y) \in A\} \geq \alpha\}$, there exists at least a sequence $\overline{x}_n$ in $\{x \in \mathbf{X} : \mathbb{P}\{(x, Y) \in A\} > \alpha\}$ converging to $\overline{x}$. It is readily seen that this corresponds to the idea underlying the second part of hypothesis (A) in [37]. Indeed, since usually the location of $\overline{x}$ is unknown, one would rather suppose that (vi) holds in order to guarantee that the second part of their hypothesis (A) holds for any possible choice of $\overline{x}$.
(iv) Hypothesis (iii) holds a fortiori if $\mathbf{X}$ is compact, as in Theorem 2.2 in [37].
(v) The stability and the accuracy of sampling in chance-constrained programs are investigated in [5, 37, 42, 43].

6 An Example

An outstanding example of min-max game or semi-infinite program is linear Chebyshev approximation. A function $\phi : Y \to \mathbb{R}$ with compact $Y \subset \mathbb{R}^m$ has to be approximated by a linear combination $\pi_N(d, y) = \sum_{j=1}^N d_j \cdot \gamma_j(y)$ of functions $\{\gamma_1, \dots, \gamma_N\}$.

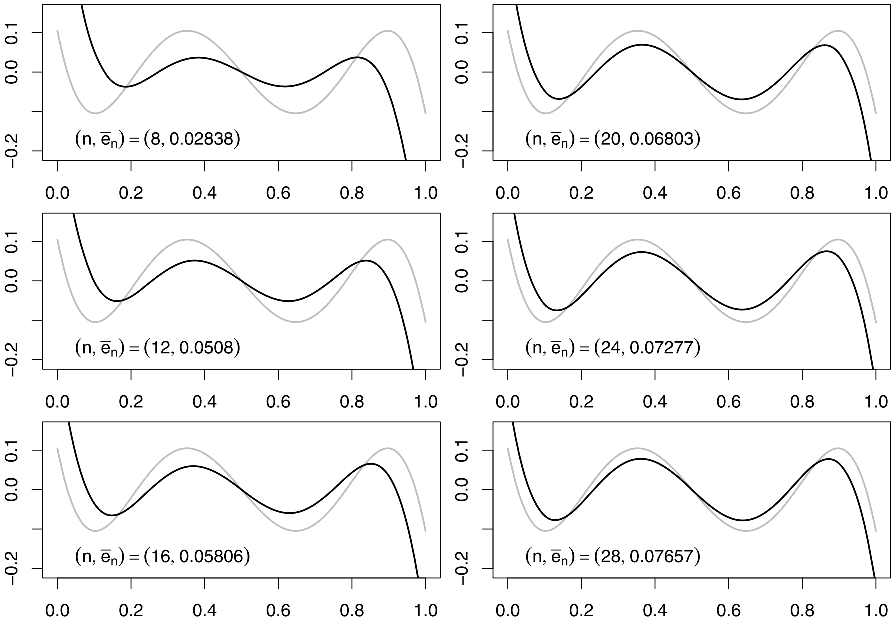

Fig. 1 Chebyshev minimizers for interior grid

The problem is the following:

$$\min_{d\in\mathbb{R}^N}\max_{y\in Y}\bigl|\phi(y)-\pi_N(d,y)\bigr|.$$

As such, it is a min-max game that can be rewritten as the semi-infinite program:

$$\min_{(d,e)\in\mathbb{R}^{N+1}} e, \quad \text{subject to} \quad \pm\bigl(\phi(y)-\pi_N(d,y)\bigr)-e\le 0, \quad \forall y\in\mathbf{Y}. \tag{13}$$

We approximate the problem using a finite collection of constraints, corresponding to the points $\{y^{(1)},\dots,y^{(n)}\}$. See [44] for more information on Chebyshev approximation and the convergence of its discretizations.

We take $N=4$, $\gamma_1(y)=1$, $\gamma_2(y)=y$, $\gamma_3(y)=2y^2-1$ and $\gamma_4(y)=4y^3-3y$, i.e. the first four Chebyshev polynomials, and $\phi(y)=\sin(2\pi y)$, defined on $Y=[0,1]$. The true coefficients, obtained through computation, are unique and are $\overline{d}=(-16.2750, 27.1599, -16.1702, 5.3901)$ and $\overline{e}=0.10473$. Taking $n\in\{8,12,16,20,24,28\}$, we have computed the minimizers of program (13) using the Barrodale and Philips algorithm (see [45]).

In the following numerical results, we consider both the case in which the points $\{y^{(1)},\dots,y^{(n)}\}$ are a subset of a sequence (the one covered by our results under the notation $(y^{(i)})_i$) and the case in which the n points constitute a point set. Clearly, any point set can be embedded in an infinite sequence, so that the previous results still apply. We consider the following cases:

1. Interior grid: It is a point set whose ith point is given by $i/(n+1)$, for $i=1,\dots,n$. This implies that the points 0 and 1 are not included in the set.

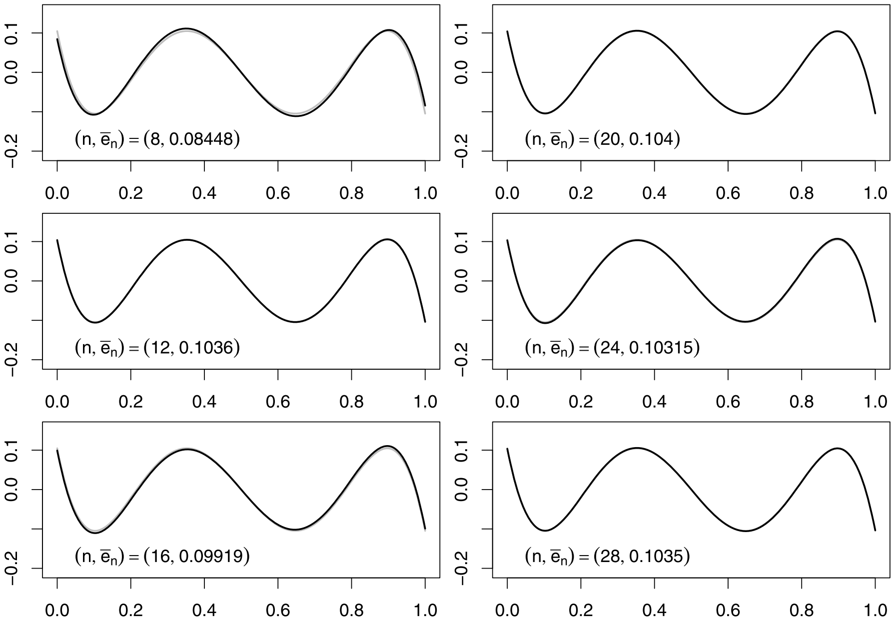

Fig. 2 Chebyshev minimizers for grid with boundary points

2. Grid with boundary points: It is a point set whose ith point is given by $(i-1)/(n-1)$, for $i = 1, \dots, n$. In this case, the boundary points 0 and 1 are included.
3. Halton sequence in base 2: It is a sequence of quasi-Monte Carlo points (see [13, p. 29]). We have chosen the base 2 since it is well known that the uniformity properties of the Halton sequence deteriorate for larger values of the base (see [13, Theorem 3.8]).
4. Random sequence: It is a sequence of points drawn according to a uniform distribution in the unit interval. In order to increase the comparability between successive plots, the points are increasing subsets (from 1 to n for every n) of the same sequence, and not separate point sets.

The first figures show the function $\phi(y) - \pi_4(\overline{d}, y)$ (in gray) together with the function $\phi(y) - \pi_4(\overline{d}_n, y)$ (in black) and display the couples of values n and $\overline{e}_n$. Figure 1 shows the minimizers in the case of the interior grid. According to [1], we expect that the rate of convergence is $O(n^{-1})$. Figure 2 shows the minimizers in the case of the grid with boundary points. The results of Still yield the rate of convergence $O(n^{-2})$. That the rate of convergence is much faster is confirmed by visual inspection, since almost no departure of the scenario minimizer with respect to the true one is visible for larger values of n. Figure 3 shows what happens when the points are randomly selected in the interval. In this case, the rate of convergence along a particular sequence is $O(\max_i \min_j |x_i - x_j|)$: using the theory of uniform spacings (see, e.g., [40, Chap. 21]), this is $O_{\mathbb{P}}(\frac{\ln n}{n})$ (Lemma 2.5 in [46]) and $O_{\mathrm{as}}(\frac{\ln n}{n})$ (Theorem 5.1 in [46]).

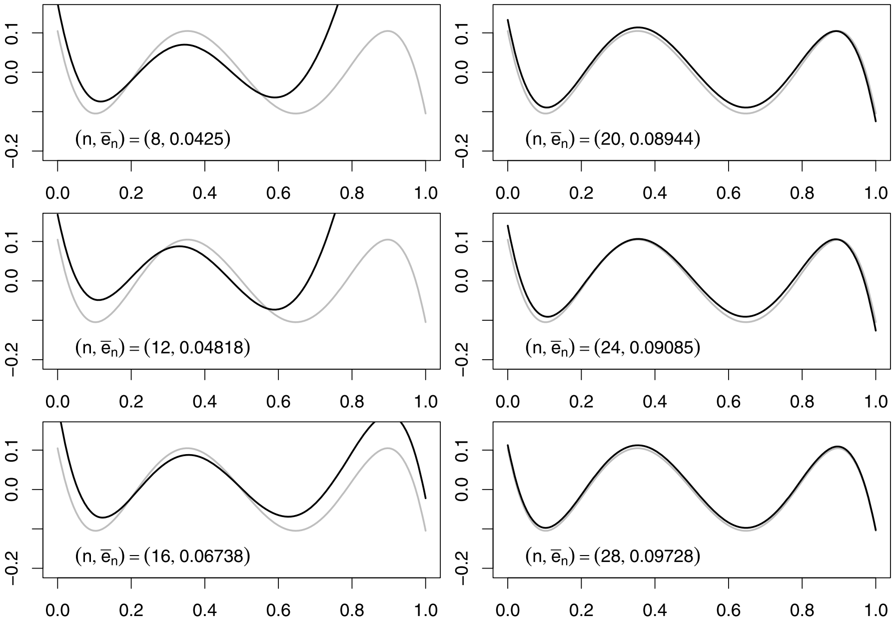

Fig. 3 Chebyshev minimizers for random sequence

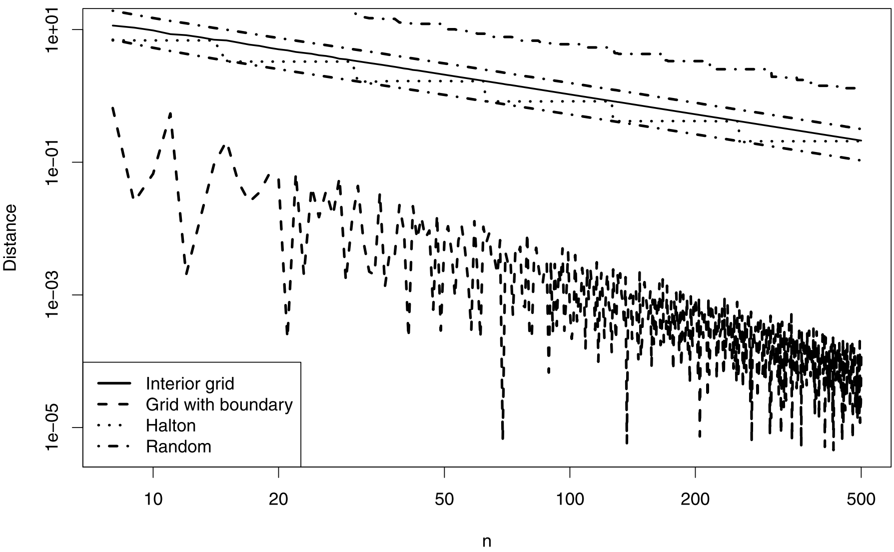

Fig. 4 Convergence rate of the parameters

Figure 4 displays the Euclidean distance between the values that minimize the scenario program and those that minimize the original program (in our notation: $\|\overline{x}_n - \overline{x}\|$, where $\overline{x}_n = (\overline{d}'_n, \overline{e}_n)'$ and $\overline{x} = (\overline{d}', \overline{e})'$) as a function of the number of

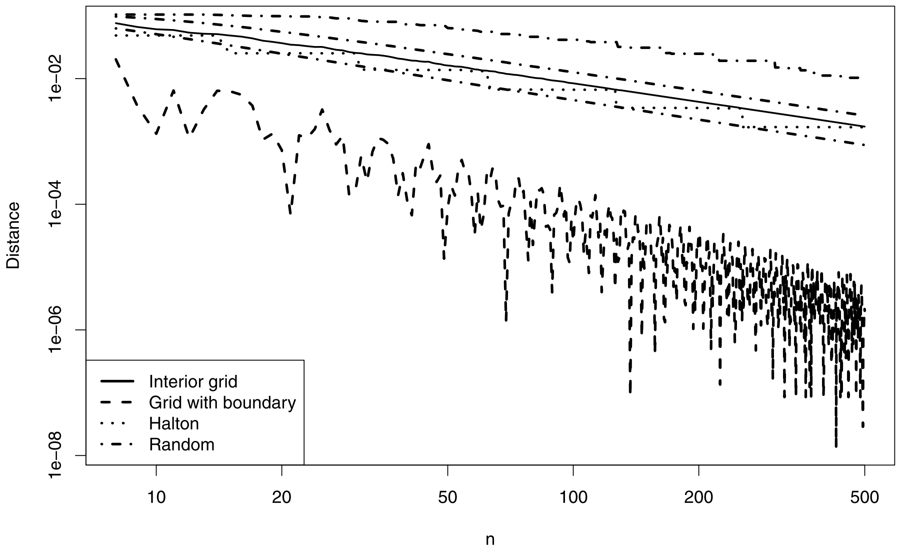

Fig. 5 Convergence rate of the objective function

points n, while Fig. 5 displays the difference of the objective functions (in this case, $|\overline{e}_n - \overline{e}|$). The continuous and long dashed lines are respectively obtained in the interior grid case and in the grid with boundary points case, and are in line with the expected convergence rate (i.e. respectively n^{-1} and n^{-2}). The oscillations arising in the second case are only in small part due to the tolerance of the optimization routine, as can be seen from the fact that they already arise for small values of n. The short dashed line is obtained when taking the Halton sequence. The particular sawtooth profile of the graph is due to the fact this sequence increases inserting regularly new points among the previous ones: these new points do not always alter the solution of the discretized program. The same sawtooth profile arises also for random sequences but in this case it is far from regular. At last, the dash-dot lines represent, starting from below, the mean, the 95 % quantile and the maximum based on 100,000 replications of random sequences. In this case, the rate of convergence appears to be slightly faster that what the reasoning above would suggest (indeed, it seems to be n^{-1}, and not $n^{-1}\ln n$).

7 Concluding Remarks

In this paper, we use epigraphical convergence to provide general conditions under which some programming problems subject to constraints (in particular min-max games, and semi-infinite, robust and chance-constrained programs) can be approximated by scenario programs. We also reformulate the definition of robust program commonly found in the literature. We illustrate our approach with a numerical example about linear Chebyshev approximation seen as a min-max game or a semi-infinite program.

Acknowledgements We are grateful to Christian Hess and Enrico Miglierina for useful comments and discussions, and to the anonymous referees, the Associate Editor Masao Fukushima and the Editor Franco Giannessi for valuable suggestions that helped improve the article substantially.

Appendix A: Proofs

Proof of Theorem 5.1 In order to show convergence of the solution of (9), we write in a different way the program. Using the indicator function χ, (9) becomes:

$$\min_{x\in\mathbf{X}\subseteq\mathbb{R}^p} c^{\mathrm{T}}x + \chi\left(x, \bigcap_{i=1}^{n}\left\{f\left(x, y^{(i)}\right) \le 0\right\}\right). \tag{14}$$

We set:

$$\begin{aligned} F_n(x) &:= \chi\left(x, \bigcap_{i=1}^{n}\left\{f\left(x, y^{(i)}\right) \le 0\right\}\right), \\ F(x) &:= \chi\left(x, \bigcap_{y\in\mathbf{Y}^\star}\left\{f(x, y) \le 0\right\}\right), \\ A(y) &:= \left\{x \in X : f(x, y) \le 0\right\}. \end{aligned}$$

To ease notation, we define $A(y) := \{x \in \mathbf{X} : f(x, y) \le 0\}$. Now, in order to show that the solution of (9) converges to the solution of (11), we use Theorem 7.33 of [33, pp. 266–267], reproduced in the Appendix as Theorem B.2. We have to verify the following hypotheses (see Definition 4.1):

1. $(F_n(x) + c^{\mathrm{T}}x)_n$ and $F(x) + c^{\mathrm{T}}x$ are lower semi-continuous and proper;
2. $(F_n(x) + c^{\mathrm{T}}x)_n$ is eventually level-bounded;
3. $(F_n(x) + c^{\mathrm{T}}x)_n$ epi-converges to $F(x) + c^{\mathrm{T}}x$.

$(F_n(x)+c^{\mathrm{T}}x)_n$ and $F(x)+c^{\mathrm{T}}x$ are lower semi-continuous: $c^{\mathrm{T}}x$ is continuous; $F_n(x)$ is lower semi-continuous iff the set $\bigcap_{i=1}^{n} A(y^{(i)})$ is closed (Example 1.6 in [30], p. 10) and this is guaranteed by f being lower semi-continuous in x for any $y \in \mathbf{Y}$ (Proposition 1.7 in [30, p. 11]). Moreover, these functions are proper as the set $\bigcap_{y\in\mathbf{Y}} A(y)$ is nonempty, and therefore:

$$\begin{aligned} F_n(x) + c^{\mathrm{T}}x &\le F(x) + c^{\mathrm{T}}x \\ &\le \chi\left(x, \bigcap_{y\in\mathbf{Y}} A(y)\right) + c^{\mathrm{T}}x \ne +\infty. \end{aligned}$$

As concerns eventual level-boundedness of the sequence $(F_n(x) + c^{\mathrm{T}}x)_n$, since the function $c^{\mathrm{T}}x$ is not level-bounded, we need the sequence of indicator functions $(F_n(x))_n$ to be eventually level-bounded: this is guaranteed by the assumption that there exists an index n_0 such that the set $\bigcap_{i=1}^{n} A(y^{(i)})$ is compact for any $n \ge n_0$.

As concerns epi-convergence, using Example 6.24(b) in [30, p. 64], we see that if $(F_n(x))_n$ epi-converges to $F(x)$ and it is an increasing sequence, and $c^{\mathrm{T}}x$ is continuous and therefore lower semi-continuous, $(F_n(x)+c^{\mathrm{T}}x)_n$ epi-converges to $F(x)+c^{\mathrm{T}}x$: all these conditions are verified apart from the epi-convergence of $(F_n(x))_n$ to $F(x)$ that has still to be proved.

According to Proposition 4.15 in [30, p. 43], epi-convergence of the indicator functions is equivalent to Painlevé–Kuratowski convergence of the sets. Since the sequence $\bigcap_{i=1}^n A(y^{(i)})$ is decreasing, according to Exercise 4.3(b) in [33, p. 111], we have:

$$\mathrm{PK}-\lim_{n\to\infty}\bigcap_{i=1}^n A\bigl(y^{(i)}\bigr)=\bigcap_{n\in\mathbb{N}}\overline{\bigcap_{i=1}^n A\bigl(y^{(i)}\bigr)}=\bigcap_{i\in\mathbb{N}} A\bigl(y^{(i)}\bigr).$$

□

Proof of Corollary 5.1 In the proof, we will use the following characterizations of lower semi-continuity of f and density of $\mathbf{Y}^\star$ in $\mathbf{Y}$. The function f is lower semi-continuous with respect to y, iff, for x fixed, we have that for every $\lambda\in\mathbb{R}$, the set $G_\lambda(x):=\{y\in\mathbf{Y}: f(x,y)>\lambda\}$ is open (see [47, Chap. 2, p. 42]); $\mathbf{Y}^\star$ is dense in $\mathbf{Y}$ iff, for all open subset G of $\mathbf{Y}$, we have $G\cap\mathbf{Y}^\star\neq\varnothing$ (see [47, Chap. 2, p. 26]).

As before, define $A(y):=\{x\in\mathbf{X}: f(x,y)\le 0\}$ and

$$I:=\bigcap_{y\in\mathbf{Y}}A(y)\quad\text{and}\quad I^\star:=\bigcap_{y^\star\in\mathbf{Y}^\star}A\bigl(y^\star\bigr).$$

To prove $I=I^\star$, we prove first $I\subseteq I^\star$ and then $I^\star\subseteq I$. The first inclusion is trivially verified since $\mathbf{Y}^\star\subseteq\mathbf{Y}$. Let us now prove that $I^\star\subseteq I$ by counterposition. Suppose that there exists an $x^\star$ such that $x^\star\in I^\star$ and $x^\star\notin I$. It respectively means that

$$\forall y^\star\in\mathbf{Y}^\star,\quad f\bigl(x^\star,y^\star\bigr)\le 0, \tag{15}$$

and $\exists y_0\in Y$ such that $f(x^\star,y_0)>0$. By combination of these two conditions, we get: $\exists y_0\in Y\backslash Y^\star$ such that $f(x^\star,y_0)>0$. Since f is lsc with respect to y, the set $G_0(x)$ is open: thus, there exists $\eta=\eta(y_0)>0$ such that, for every $y\in\mathsf{B}(y_0,\eta)$ (the ball of radius η centered in y_0), $y\in G_0$ i.e. $f(x,y)>0$. But, since $\mathbf{Y}^\star$ is dense in $\mathbf{Y}$, there exists some $y_0^\star\in\mathsf{B}(y_0,\eta)\cap\mathbf{Y}^\star$. So $f(x^\star,y_0^\star)>0$, which contradicts (15). Therefore:

$$\forall y\in\mathbf{Y}\setminus\mathbf{Y}^\star,\quad f\bigl(x^\star,y\bigr)\le 0,$$

which, together with (15), gives

$$\forall y\in\mathbf{Y},\quad f\bigl(x^\star,y\bigr)\le 0.$$

Thus, we have $x^\star\in I$ and $I^\star\subseteq I$. As a consequence, $I=I^\star$. □

Proof of Corollary 5.2 As in the previous theorem, we just need to show that $(F_n(x))_n$ epi-converges to a certain limit function $F(x)$, i.e. that

$$\bigcap_{i=1}^{n} \{ f(x, y^{(i)}) \le 0 \}$$

converges in the sense of Painlevé–Kuratowski to a limit set (Proposition 4.15 in [30, p. 43]). Using Theorem 2.7 in [28], we see that $\bigcap_{i=1}^{n}\{f(x, y^{(i)}) \le 0\}$ converges in the sense of Painlevé–Kuratowski to $\mathrm{Int}(\{x \in \mathbf{X} : f(x, Y) \le 0\})$ under the following conditions:

(i) The set $\{f(x, y^{(i)}) \le 0\}$ has to be nonempty and closed: nonemptiness is guaranteed by the statement of the corollary, and closedness by the fact that f is lower semi-continuous in x for $\mathbb{P}$-almost surely any $y \in \mathbf{Y}$ (Proposition 1.7 in [30, p. 17]).

(ii) The random variable $D(Y) := d(0, \mathrm{Int}(\{x \in \mathbf{X} : f(x, Y) \le 0\}))$ is integrable.

Therefore, $(F_n(x))_n$ epi-converges $\mathbb{P}$-as to

$$F(x) = \chi\big(x, \mathrm{Int}\big(\big\{x \in \mathbf{X} : f(x, Y) \le 0\big\}\big)\big).$$

Note that we can use a result of [18, Proposition 21, p. IV-34] to write

$$\begin{aligned} F(x) &= \chi\big(x, \mathrm{Int}\big\{x \in \mathbf{X} : f(x, Y) \le 0\big\}\big) \\ &= \int_{\mathbf{Y}} \chi\big(x, \big\{x \in \mathbf{X} : f(x, y) \le 0\big\}\big)\mathbb{P}(dy), \end{aligned}$$

and to express (8) as

$$\min_{x \in \mathbf{X} \subseteq \mathbb{R}^n} c^{\mathrm{T}} x + \int_{\mathbf{Y}} \chi\big(x, \big\{x \in \mathbf{X} : f(x, y) \le 0\big\}\big)\mathbb{P}(dy).$$

Convergence of the solution can be proved using the same conditions as before (see the proof of Theorem 5.1): in particular, we have just to check for eventual level-boundedness of $(F_n(x))_n$, but this is guaranteed by integrability of $d(0, \{f(x, Y) \le 0\})$. □

Proof of Theorem 5.2 The idea is to write this program, using the properties of the indicator function χ, as $\min_{x \in \mathbf{X} \subseteq \mathbb{R}^p} a(x)$, where:

$$a(x) = h(x) + \chi\big(x, \mathrm{lev}_{\le -\alpha} \mathbb{E}\big[-\mathbf{1}\big\{(x, Y) \in A\big\}\big]\big).$$

The approximate solution is given by $\min_{x \in X} a_n(x)$, where:

$$a_n(x) = h(x) + \chi\left(x, \mathrm{lev}_{\le -\alpha}\left\{\frac{1}{n}\sum_{i=1}^{n}\big[-\mathbf{1}\big\{\big(x, y^{(i)}\big) \in A\big\}\big]\right\}\right).$$

If we define the empirical probability based on the sequence $(y^{(i)})_{i=1,\dots,n}$ as $\mathbb{P}_n(B) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(y^{(i)} \in B)$, we have

$$\mathbb{P}_n\big\{(x, Y) \in A\big\} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\big\{\big(x, y^{(i)}\big) \in A\big\}.$$

(a) Under hypothesis (i), the function $-\mathbf{1}\{(x, Y) \in A\}$ is lower semi-continuous in x for $\mathbb{P}$-almost any $Y \in \mathbf{Y}$ and measurable with respect to $\mathcal{B}(\mathbf{X}) \otimes \mathcal{Y}$. We can then apply Corollary 2.4 in [28, p. 70], to prove that, for almost any independent and identically distributed sequence $(y^{(i)})_{i=1,\dots,n}$,

$$\begin{aligned} \mathrm{epi} - \lim_{n\to\infty} -\mathbb{P}_n\{(x, y^{(i)}) \in A\} &= \mathrm{epi} - \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left[-\mathbf{1}\{(x, y^{(i)}) \in A\}\right] \\ &= -\mathbb{P}\{(x, Y) \in A\} \quad \mathbb{P}\text{-as.} \end{aligned} \tag{16}$$

Now, define

$$\begin{aligned} S_n\big(\alpha, (y^{(i)})_i\big) &:= \mathrm{lev}_{\le -\alpha}\{-\mathbb{P}_n\{(x, y^{(i)}) \in A\}\}, \\ S(\alpha) &:= \mathrm{lev}_{\le -\alpha}\{-\mathbb{P}\{(x, Y) \in A\}\}, \end{aligned}$$

for $\alpha \in \mathbb{R}$. Using the characterization of epi-convergence through level sets (see [48, result (b) on p. 755]),

$$\mathrm{PK} - \limsup_{n\to\infty} S_n\big(\alpha_n, (y^{(i)})_i\big) \subset S(\alpha) \quad \mathbb{P}\text{-as,} \tag{17}$$

for any $(\alpha_n)_n$ such that $\alpha_n \to \alpha$. Recall that, for a sequence of sets $(C_n)_n$, $\mathrm{PK} - \limsup_{n\to\infty} C_n$ is the set of all cluster points extracted from the sequence $(C_n)_n$: this means that if we define a sequence $(\overline{x}_n)_n$ through $\overline{x}_n \in \arg\min_{x\in\mathbf{X}} a_n(x)$, the set of cluster points of $(\overline{x}_n)_n$ is included in $S(\alpha)$.

From Proposition 4.15 in [30, p. 43], Eq. (17) means that

$$\mathrm{epi} - \liminf_{n\to\infty} \chi\big(x, S_n\big(\alpha_n, (y^{(i)})_i\big)\big) \ge \chi\big(x, S(\alpha)\big) \quad \mathbb{P}\text{-as.}$$

From Proposition 6.21 in [30, p. 63], since $h(x)$ is continuous by (ii),

$$\mathrm{epi} - \liminf_{n\to\infty} a_n(x) = h(x) + \mathrm{epi} - \liminf_{n\to\infty} \chi\big(x, S_n\big(\alpha_n, (y^{(i)})_i\big)\big) \ge a(x)\ \mathbb{P}\text{-as.}$$

Proposition 7.29 (a) in [33] yields $\liminf_n a_n(\overline{x}_n) \ge a(\overline{x})$ provided (iii) holds, and from the obvious relations $a_n(\overline{x}_n) = h(\overline{x}_n)$ and $a(\overline{x}) = h(\overline{x})$, we get the desired result.

(b) From (16), using Proposition 7.7 (b) in [33], we obtain the existence of a sequence $(\alpha_n^\star)_n$ with $\alpha_n^\star \uparrow \alpha$ such that

$$S(\alpha) \subseteq \liminf_{n\to\infty} S_n\big(\alpha_n^\star, (y^{(i)})_i\big) \quad \mathbb{P}\text{-as.}$$

Therefore we get:

$$\begin{aligned} &\mathrm{epi} - \liminf_{n\to\infty} \left[h(x) + \chi\big(x, S_n\big(\alpha_n^\star, (y^{(i)})_i\big)\big)\right] \\ &= h(x) + \mathrm{epi} - \liminf_{n\to\infty} \chi\big(x, S_n\big(\alpha_n^\star, (y^{(i)})_i\big)\big) \\ &= h(x) + \chi\big(x, S(\alpha)\big) \quad \mathbb{P}\text{-as,} \end{aligned}$$

where the first equality derives from Proposition 6.21 in [30, p. 63], since $h(x)$ is continuous by (ii), and the second one from the property of the particular sequence $(\alpha_n^\star)_n$.

Now, we show convergence of the minimizers. The functions $a(x)$ and $a_n(x)$ are lower semi-continuous since, under hypothesis (i), the function $\mathbf{1}\{(x,Y)\in A\}$ is lower semi-continuous in x for $\mathbb{P}$-almost any $Y \in \mathbf{Y}$. Under hypotheses (ii), (iii) and (iv), $a_n(x)$ and $a(x)$ are proper. The sequence $(a_n(x))_n$ is eventually level-bounded, from (ii) and (iii). Therefore, for the sequence $(\alpha_n^\star)_n$, Theorem B.2 applies.

(c) Then, we pass to the last part of the theorem. We start from the first statement. In particular, the fact that $\liminf_n h(\overline{x}_n) \geq h(\overline{x})$ $\mathbb{P}$-almost surely is a consequence of (a) under hypotheses (i)–(ii)–(iii)–(iv) with $\alpha_n = \alpha$. As concerns the fact that $\limsup_n h(\overline{x}_n) \leq h(\overline{x})$ $\mathbb{P}$-almost surely, it can be shown to hold following the proof of Proposition 2.2 in [37] and replacing the fact that G is Carathéodory with (i), continuity of f with (ii), compactness of X with (iii), the existence of an optimal solution stated in (A) with (iv)–(ii) and the remaining part of (A) with (vi) (see Remark 5.3 for a comparison of the hypotheses). As concerns the second statement, we first show that $S_n(\alpha, (y^{(i)})_i)$ converges $\mathbb{P}$-almost surely in the Hausdorff metric to $S(\alpha)$, then we show that this implies epi-convergence of the objective functions and we close the proof proving convergence of the solutions. Let $A(x)$ be the set defined in the statement of the theorem. Then we have $\mathbf{1}\{(x,Y)\in A\} = \mathbf{1}\{Y \in A(x)\}$ and:

$$\begin{aligned}&\sup_{x\in\mathbf{X}}\bigl|\mathbb{P}_n\bigl\{(x,Y)\in A\bigr\} - \mathbb{P}\bigl\{(x,Y)\in A\bigr\}\bigr|\\&\quad = \sup_{x\in\mathbf{X}}\bigl|\mathbb{P}_n\bigl\{Y\in A(x)\bigr\} - \mathbb{P}\bigl\{Y\in A(x)\bigr\}\bigr| \underset{n\to\infty}{\longrightarrow} 0 \quad \mathbb{P}\text{-as.}\end{aligned}$$

This holds because of hypothesis (v). On the other hand, Eq. (2.3) in [49] becomes

$$\bigl\{x\in\mathbf{X} : \mathbb{P}\bigl\{Y\in A(x)\bigr\}\geq\alpha\bigr\} \subseteq \mathrm{cl}\bigl[\bigl\{x\in\mathbf{X} : \mathbb{P}\bigl\{Y\in A(x)\bigr\}>\alpha\bigr\}\bigr],$$

and this is equivalent to hypothesis (vi). From hypothesis (iii), $S_n(\alpha, (y^{(i)})_i)$ $\mathbb{P}$-almost surely converges in the Hausdorff metric to $S(\alpha)$ by Theorem 2.1 in [49].

On the space of nonempty compact subsets of an Euclidean space, convergence in the Hausdorff metric and Painlevé–Kuratowski convergence of sequences of sets are equivalent, and both are equivalent to epi-convergence of the indicator functions of the sets (Proposition 4.15 in [30, p. 43]). This means that $\mathrm{PK} - \lim_{n\to\infty} S_n(\alpha, (y^{(i)})_i) = S(\alpha)$ $\mathbb{P}$-as is equivalent to:

$$\mathrm{epi} - \lim_{n\to\infty} \chi\bigl(x, S_n\bigl(\alpha, \bigl(y^{(i)}\bigr)_i\bigr)\bigr) = \chi\bigl(x, S(\alpha)\bigr) \quad \mathbb{P}\text{-as.}$$

From Proposition 6.21 in [30], since $h(x)$ is continuous by (ii),

$$\mathrm{epi} - \lim_{n\to\infty} h(x) + \chi\bigl(x, S_n\bigl(\alpha, \bigl(y^{(i)}\bigr)_i\bigr)\bigr) = h(x) + \chi\bigl(x, S(\alpha)\bigr) = a(x) \quad \mathbb{P}\text{-as.}$$

Therefore, we can apply Theorem B.2 that holds since the objective functions $(a_n(x))_n$ and $a(x)$ are lower semi-continuous, proper and eventually level-bounded (from (ii) and (iii)), $(a_n(x))_n$ is epi-convergent to $a(x)$ and the space $\mathbf{X}$ is compact. □

Appendix B: Some Mathematical Concepts

B.1 Epi-convergence

Since our main result is based on epi-convergence, we provide a short presentation. Let $h : E \to \overline{\mathbb{R}}$ be a function from the metric space E into the extended reals. Its *epigraph* is defined by

$$\mathrm{Epi}(h) := \big\{(x,\lambda) \in E \times \mathbb{R} : h(x) \le \lambda\big\}.$$

The *hypograph* of h, denoted by $\mathrm{Hypo}(h)$, is defined by reversing the inequality. Let $(h_n)_{n\ge 1}$ (or $(h_n)_n$ for short) be a sequence of functions from E into $\overline{\mathbb{R}}$. For any $x \in E$, we introduce the quantities

$$\begin{aligned} \mathrm{epi} - \liminf_{n\to\infty} h_n(x) &:= \sup_{k\ge 1} \liminf_{n\to\infty} \inf_{y\in \mathsf{B}(x,1/k)} h_n(y), \\ \mathrm{epi} - \limsup_{n\to\infty} h_n(x) &:= \sup_{k\ge 1} \limsup_{n\to\infty} \inf_{y\in \mathsf{B}(x,1/k)} h_n(y), \end{aligned} \tag{18}$$

where $\mathsf{B}(x,1/k)$ denotes the open ball of radius $1/k$ centered at x. The function $x \mapsto \mathrm{epi} - \liminf_{n\to\infty} h_n(x)$ (resp. $x \mapsto \mathrm{epi} - \limsup_{n\to\infty} h_n(x)$) is called the *lower* (resp. *upper*) *epi-limit* of the sequence $(h_n)_n$. These functions are lsc. If $\mathrm{epi} - \liminf_{n\to\infty} h_n(x) = \mathrm{epi} - \limsup_{n\to\infty} h_n(x)$, then $(h_n)_n$ is said to be *epi-convergent* at x. If this is true for all $x \in E$, then the sequence $(h_n)_n$ *epi-converges*. Its *epi-limit* is denoted by $\mathrm{epi} - \lim_{n\to\infty} h_n$.

Equalities (18) have a geometric counterpart involving the Painlevé–Kuratowski convergence of epigraphs on the space of closed sets of $E \times \mathbb{R}$ (see, e.g., [50] or [30]). The Painlevé–Kuratowski convergence is defined as follows. Given a sequence $(C_n)_{n\ge 1}$ of sets in E, we define

$$\begin{aligned} \mathrm{PK} - \liminf_{n\to\infty} C_n &:= \{x \in E : x = \lim x_n,\ x_n \in C_n,\ \forall n \ge 1\}, \\ \mathrm{PK} - \limsup_{n\to\infty} C_n &:= \{x \in E : x = \lim x_i,\ x_i \in C_{n(i)},\ \forall i \ge 1\}, \end{aligned}$$

where $(C_{n(i)})_{i\ge 1}$ is a subsequence of $(C_n)_{n\ge 1}$. The subsets $\mathrm{PK} - \liminf_{n\to\infty} C_n$ and $\mathrm{PK} - \limsup_{n\to\infty} C_n$ are the *lower limit* and the *upper limit* of $(C_n)_{n\ge 1}$. It is not difficult to check that they are both closed and that they satisfy $\mathrm{PK} - \liminf_{n\to\infty} C_n \subset \mathrm{PK} - \limsup_{n\to\infty} C_n$. A sequence $(C_n)_{n\ge 1}$ is said to *converge* to C, in the sense of *Painlevé–Kuratowski*, if

$$C = \mathrm{PK} - \liminf_{n\to\infty} C_n = \mathrm{PK} - \limsup_{n\to\infty} C_n.$$

This is denoted by $C = \mathrm{PK} - \lim_{n\to\infty} C_n$. As mentioned above, this notion is strongly connected with epi-convergence: a sequence of functions $h_n : E \to \overline{\mathbb{R}}$ epi-converges to h iff the sequence $(\mathrm{Epi}(h_n))_{n\ge 1}$ PK-converges to $\mathrm{Epi}(h)$, in $E \times \mathbb{R}$.

A characterization of epi-convergence can be given using level sets (see [33, p. 246]).

Theorem B.1 *Let* $h : \mathbb{R}^d \to \mathbb{R}$ *and* $(h_n)_n$ *be such that* $h_n : \mathbb{R}^d \to \mathbb{R}$. *Then*:

(i) $\operatorname{epi} - \liminf_{n\to\infty} h_n \ge h$ *iff*

$$\limsup_n (\operatorname{lev}_{\le \alpha_n} h_n) \subseteq \operatorname{lev}_{\le \alpha} h,$$

for all sequences $\alpha_n \to \alpha$;

(ii) $h \ge \operatorname{epi} - \limsup_{n\to\infty} h_n$ *iff*

$$\liminf_n (\operatorname{lev}_{\le \alpha_n} h_n) \supseteq \operatorname{lev}_{\le \alpha} h,$$

for some sequence $\alpha_n \to \alpha$, *in which case this sequence can be chosen with* $\alpha_n \downarrow \alpha$;

(iii) $\operatorname{epi} - \lim_{n\to\infty} h_n = h$ *if and only if both conditions hold.*

B.2 Convergence of Minima

The following result (Theorem 7.33 in [33, pp. 266–267]) plays a fundamental role in our proofs.

Theorem B.2 *Suppose that the sequence* $(h_n)_n$ *is eventually level-bounded, and* $\operatorname{epi} - \lim_{n\to\infty} h_n = h$ *with* h_n *and* h *lower semi-continuous and proper. Then*:

$$\inf h_n \to \inf h$$

and $\inf h$ *is finite*; *moreover*, *there exists* n_0 *such that*, *for any* $n \ge n_0$, *the sets* $\arg\min h_n$ *are nonempty and form a bounded sequence with*

$$\limsup_n (\arg\min h_n) \subseteq \arg\min h.$$

Indeed, *for any choice of* $\varepsilon_n \downarrow 0$ *and* $x_n \in \varepsilon_n - \arg\min h_n$, *the sequence* $(x_n)_{n\in\mathbb{N}}$ *is bounded and such that all its cluster points belong to* $\arg\min h$. *If* $\arg\min h$ *consists of a unique point* $\overline{x}$, *one must actually have* $x_n \to \overline{x}$.

B.3 Stationarity and Ergodicity

A sequence of random variables $(X_i)_{i=1,\dots}$ is said to be *stationary* if the random vectors $(X_1, \dots, X_n)$ and $(X_{k+1}, \dots, X_{n+k})$ have the same distribution for all integers $n, k \ge 1$. A measurable set B is said to be *invariant* if

$$\bigl\{(X_i)_{i=1,\dots} \in B\bigr\} = \bigl\{(X_i)_{i=k,k+1,\dots} \in B\bigr\}$$

for every $k \ge 1$. A sequence $(X_i)_{i=1,\dots}$ is said to be *ergodic* if, for every invariant set B,

$$\mathbb{P}\bigl\{(X_i)_{i=1,\dots} \in B\bigr\} \in \{0, 1\}.$$

These properties can be introduced also in a more abstract setting. Given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, an $\mathcal{A}$-measurable transformation $T : \Omega \to \Omega$ is said to be *measure-preserving* if $\mathbb{P}(T^{-1}A) = \mathbb{P}(A)$ for all $A \in \mathcal{A}$. Equivalently, $\mathbb{P}$ is said to be *stationary* with respect to T. The sets $A \in \mathcal{A}$ that satisfy $T^{-1}A = A$ are called *invariant sets* and constitute a sub-σ-field $\mathcal{I}$ of $\mathcal{A}$. A measurable and measure-preserving transformation T is said to be *ergodic* if $\mathbb{P}(A) = 0$ or 1 for all invariant sets A. Equivalently, the sub-σ-field $\mathcal{I}$ reduces to the trivial σ-field $\{\Omega, \varnothing\}$ (up to the $\mathbb{P}$-null sets). The previous definitions can be recovered remarking that any stationary sequence $(X_i)_{i=1,\dots}$ can almost surely be rewritten using a measurable and measure-preserving transformation T as $X_t(\omega) = X_0(T^t\omega)$ (see, e.g., [51, Proposition 6.11]).

B.4 Random Sets

Given a Polish space E, the set of all subsets of E is denoted by 2^E. A random set is a set-valued map $\Gamma : \Omega \to 2^E$ having some sort of measurability property. Here, we shall use graph measurability. The *graph* of Γ is denoted by $Gr(\Gamma)$ and defined by

$$Gr(\Gamma) = \big\{(\omega, x) \in \Omega \times E : x \in \Gamma(\omega)\big\}.$$

In this framework, Γ is said to be a *random set* if $Gr(\Gamma)$ is a member of the product σ-field $\mathcal{A} \otimes \mathcal{B}(E)$. Then, Γ is said to be *graph-measurable.*

References

1. Still, G.: Discretization in semi-infinite programming: The rate of convergence. Math. Program., Ser. A **91**(1), 53–69 (2001)
2. Calafiore, G., Campi, M.C.: Robust convex programs: Randomized solutions and applications in control. In: Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii, USA, December 2003 pp. 2423–2428 (2003)
3. Campi, M.C., Calafiore, G.: Decision making in an uncertain environment: The scenario-based optimization approach. In: Kárný, M., Kracík, J., Andrýsek, J. (eds.) Multiple participant decision making. International Series on Advanced Intelligence, vol. 9, pp. 99–111. Advanced Knowledge International (2004)
4. Calafiore, G., Campi, M.C.: Uncertain convex programs: Randomized solutions and confidence levels. Math. Program. **102**, 25–46 (2005)
5. Nemirovski, A., Shapiro, A.: Scenario approximations of chance constraints. In: Calafiore, G., Dabbene, F. (eds.) Probabilistic and Randomized Methods for Design under Uncertainty, pp. 3–48. Springer, Berlin (2006)
6. Pucci de Farias, D., Van Roy, B.: On constraint sampling in the linear programming approach to approximate dynamic programming. Math. Oper. Res. **29**(3), 462–478 (2004)
7. Reemtsen, R.: Semi-infinite programming: Discretization methods. In: Floudas, C., Pardalos, P. (eds.) Encyclopedia of Optimization, pp. 3417–3424. Springer, New York (2009)
8. Reemtsen, R.: Discretization methods for the solution of semi-infinite programming problems. J. Optim. Theory Appl. **71**(1), 85–103 (1991)
9. Reemtsen, R.: Some outer approximation methods for semi-infinite optimization problems. J. Comput. Appl. Math. **53**(1), 87–108 (1994)
10. Combettes, P.: Strong convergence of block-iterative outer approximation methods for convex optimization. SIAM J. Control Optim. **38**(2), 538–565 (2000) (electronic)
11. Shapiro, A.: Monte Carlo sampling methods. In: Stochastic programming. Handbooks Oper. Res. Management Sci., vol. 10, pp. 353–425. Elsevier, Amsterdam (2003)
12. Huber, P.J.: The 1972 Wald lecture. Robust statistics: A review. Ann. Math. Stat. **43**, 1041–1067 (1972)
13. Niederreiter, H.: Random number generation and quasi-Monte Carlo methods. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1992)

14. Polak, E.: On the mathematical foundations of nondifferentiable optimization in engineering design. SIAM Rev. **29**(1), 21–89 (1987)
15. Hettich, R., Kortanek, K.O.: Semi-infinite programming: Theory, methods, and applications. SIAM Rev. **35**(3), 380–429 (1993)
16. Žaković, S., Rustem, B.: Semi-infinite programming and applications to minimax problems. Ann. Oper. Res. **124**, 81–110 (2003)
17. Hess, C., Seri, R., Choirat, C.: Approximations results for robust optimization. Working paper (2010)
18. Hiriart-Urruty, J.B.: Contributions à la programmation mathématique: cas déterministe et stochastique. Ph.D. thesis, Université de Clermont-Ferrand II, Clermont (1977)
19. Charnes, A., Cooper, W.W., Symonds, G.H.: Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. Manag. Sci. **4**, 183–195 (1958)
20. Charnes, A., Cooper, W.W.: Chance-constrained programming. Manag. Sci. **6**, 73–79 (1959/1960)
21. Charnes, A., Cooper, W.W.: Chance constraints and normal deviates. J. Am. Stat. Assoc. **57**, 134–148 (1962)
22. Sengupta, J.K.: Stochastic linear programming with chance constraints. Int. Econ. Rev. **11**, 101–116 (1970)
23. Still, G.: Generalized semi-infinite programming: Theory, methods. Eur. J. Oper. Res. **119**, 301–313 (1999)
24. Still, G.: Generalized semi-infinite programming: Numerical aspects. Optimization **49**(3), 223–242 (2001)
25. Bai, D., Carpenter, T.J., Mulvey, J.M.: Making a case for robust models. Manag. Sci. **43**, 895–907 (1997)
26. Redaelli, G.: Convergence problems in stochastic programming models with probabilistic constraints. Riv. Mat. Sci. Econ. Soc. **21**(1–2), 147–164 (1998)
27. Pennanen, T., Koivu, M.: Epi-convergent discretizations of stochastic programs via integration quadratures. Numer. Math. **100**, 141–163 (2005)
28. Choirat, C., Hess, C., Seri, R.: A functional version of the Birkhoff ergodic theorem for a normal integrand: A variational approach. Ann. Probab. **31**(1), 63–92 (2003)
29. Choirat, C., Hess, C., Seri, R.: Approximation of stochastic programming problems. In: Niederreiter, H., Talay, D. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2004, pp. 45–60. Springer, Berlin (2006)
30. Dal Maso, G.: An Introduction to Γ-Convergence. Progress in Nonlinear Differential Equations and their Applications, vol. 8. Birkhäuser, Boston (1993)
31. Jagannathan, R.: Chance-constrained programming with joint constraints. Oper. Res. **22**(2), 358–372 (1974)
32. López, M., Still, G.: Semi-infinite programming. Eur. J. Oper. Res. **180**(2), 491–518 (2007)
33. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 317. Springer, Berlin (1998)
34. Allen, F.M., Braswell, R.N., Rao, P.V.: Distribution-free approximations for chance constraints. Oper. Res. **22**(3), 610–621 (1974)
35. Gray, R.M., Kieffer, J.C.: Asymptotically mean stationary measures. Ann. Probab. **8**(5), 962–973 (1980)
36. Wets, R.J.B.: Stochastic programs with chance constraints: Generalized convexity and approximation issues. In: Generalized Convexity, Generalized Monotonicity: Recent Results, Luminy, 1996. Nonconvex Optim. Appl., vol. 27, pp. 61–74. Kluwer Academic, Dordrecht (1998)
37. Pagnoncelli, B.K., Ahmed, S., Shapiro, A.: Sample average approximation method for chance constrained programming: Theory and applications. J. Optim. Theory Appl. **142**(2), 399–416 (2009)
38. Elker, J., Pollard, D., Stute, W.: Glivenko–Cantelli theorems for classes of convex sets. Adv. Appl. Probab. **11**(4), 820–833 (1979)
39. Steele, J.M.: Empirical discrepancies and subadditive processes. Ann. Probab. **6**(1), 118–127 (1978)
40. Shorack, G.R., Wellner, J.A.: Empirical Processes with Applications to Statistics. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, New York (1986)
41. Nobel, A.: A counterexample concerning uniform ergodic theorems for a class of functions. Stat. Probab. Lett. **24**(2), 165–168 (1995)
42. Henrion, R., Römisch, W.: Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. Math. Program., Ser. A **84**(1), 55–88 (1999)
43. Luedtke, J., Ahmed, S.: A sample approximation approach for optimization with probabilistic constraints. SIAM J. Optim. **19**(2), 674–699 (2008)

44. Cheney, E.W.: Introduction to Approximation Theory. AMS Chelsea, Providence (1998). Reprint of the second (1982) edition
45. Barrodale, I., Phillips, C.: Algorithm 495: Solution of an overdetermined system of linear equations in the Chebyshev norm. ACM Trans. Math. Softw. **1**(3), 264–270 (1975)
46. Devroye, L.: Laws of the iterated logarithm for order statistics of uniform spacings. Ann. Probab. **9**(5), 860–867 (1981)
47. Aliprantis, C.D., Border, K.C.: Infinite-Dimensional Analysis. Springer, Berlin (1999)
48. Beer, G., Rockafellar, R.T., Wets, R.J.B.: A characterization of epi-convergence in terms of convergence of level sets. Proc. Am. Math. Soc. **116**(3), 753–761 (1992)
49. Molchanov, I.S.: A limit theorem for solutions of inequalities. Scand. J. Stat. **25**(1), 235–242 (1998)
50. Attouch, H.: Variational Convergence for Functions and Operators. Applicable Mathematics Series. Pitman, Boston (1984)
51. Breiman, L.: Probability. Classics in Applied Mathematics, vol. 7. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1992)