SIAM J. OPTIM.
Vol. 28, No. 1, pp. 135–162

CONSISTENCY OF THE SCENARIO APPROACH*

FEDERICO ALESSANDRO RAMPONI†

I dedicate this work to the memory of my father Alessandro Ramponi, 1945–2010
Cuius nomen iam demum in meo resonat, ille rector et una mecum auctor

Abstract. This paper is meant to prove the consistency of the scenario approach à la Calafiore, Campi, and Garatti with convex constraints. Scenario convex problems are usually stated in two equivalent forms: first, as the minimum of a linear function over the intersection of a finite random sample of independent and identically distributed convex sets, or second, as the min-max of a finite random sample of independent and identically distributed convex functions. The paper shows that, under fairly general assumptions, as the size of the sample increases the minimum attained by the solution of a problem of the first kind converges almost surely to the minimum attained by a suitably defined "essential" robust problem (or diverges if such a robust problem is infeasible), and that the minimum attained by the solution of a scenario problem of the second kind converges almost surely to the minimum of the pointwise essential supremum taken over all the possible convex functions (or diverges if such essential supremum takes the only value $+\infty$). In both cases, if the solution of the essential problem exists and is unique, the solution of the scenario problem converges to it almost surely.

Key words. scenario approach, stochastic programming, convex programming, sample-based optimization, min-max optimization

AMS subject classifications. 90C15, 90C25, 90C47

DOI. 10.1137/16M109819X

1. Introduction.

Robust convex optimization deals with programs of the following form:

$$\begin{aligned} &\min_{\theta\in\Theta}\; c^\top \theta \\ &\text{subject to} \quad \theta \in \Theta_\delta \text{ for all } \delta \in \Delta, \end{aligned} \tag{1}$$

where Θ is a closed convex subset of $\mathbb{R}^d$, Δ is an arbitrary set (possibly infinite), and to every $\delta \in \Delta$ there is associated a closed convex subset $\Theta_\delta \subseteq \Theta$. Any Θ_δ, $\delta \in \Delta$, is interpreted as an additional constraint to the nominal problem $\min_{\theta\in\Theta}\; c^\top\theta$, i.e., a constraint that may appear in a practical instance but whose actual occurrence is not known in advance; the solution of (1) is thus a safeguard against all the possible deviations from the nominal problem.

In practical applications dealing with *all* the possible constraints Θ_δ, $\delta \in \Delta$, may be overkill, and discarding a small fraction of constraints, i.e., a small subset of Δ, is often acceptable. The set Δ models the lack of knowledge in an optimization endeavor, and in science and engineering the natural way to model uncertainty is through probability; thus, from now on, I will assume that $(\Delta, \mathcal{F}, \mathsf{P})$ is a probability space, where P describes the chance of a constraint set Θ_δ to occur. Moreover, $(\Delta^N, \mathcal{F}^N, \mathsf{P}^N)$ will denote the N-fold Cartesian product of Δ equipped with the product σ-algebra $\mathcal{F}^N$ and the product probability $\mathsf{P}^N = \mathsf{P} \times \cdots \times \mathsf{P}$ (N times). A point in $(\Delta^N, \mathcal{F}^N, \mathsf{P}^N)$ will thus be a sample $(\delta^{(1)}, \ldots, \delta^{(N)})$ of elements drawn independently from Δ ac-

*Received by the editors October 10, 2016; accepted for publication (in revised form) October 9, 2017; published electronically January 11, 2018.
http://www.siam.org/journals/siopt/28-1/M109819.html

†Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Brescia, 25123 Brescia, Italy (federico.ramponi@unibs.it).

cording to the same probability P. One possible approach to weaken problem (1) by accepting a small portion of constraints, i.e., a subset of Δ with small probability ε, to be violated, is the following so-called chance-constrained problem:

$$\begin{aligned} &\min_{\theta \in \Theta} \; c^\top \theta \\ &\text{subject to} \quad \mathsf{P}\left[\{\delta \in \Delta \;:\; \theta \in \Theta_\delta\}\right] \geq 1 - \varepsilon. \end{aligned} \tag{2}$$

Let me provide a visual explanation of the difference between robust and chance-constrained problems with a toy example.

Example 1. Suppose that $\Delta = [-1, 1]$, and consider the problem

$$\begin{aligned} &\min_{(x,y) \in \mathbb{R}^2} \; x + y \\ &\text{subject to} \quad (x - \delta)^2 + y^2 \leq 4 \text{ for all } \delta \in \Delta. \end{aligned} \tag{3}$$

Clearly, (3) is an instance of problem (1), where $\theta = (x, y)$ and Θ_δ is the closed ball with center $(\delta, 0)$ and radius 2. A pictorial view of the problem is shown in Figure 1(a), where the set $\Delta \times \{0\}$ containing the center of each closed ball is the thick line segment at the center of the plot. The whole point of robust programming is that each ball may be the "true one" that will show up in reality; some balls represent favorable situations (the leftmost ones) and other bad situations, but in order to take into account *all* the constraint sets, feasible points are a priori confined to their intersection: the feasible set of problem (3) is indeed the white oval at the center of the plot, and its solution is marked with a bullet (•). Suppose now that $\Delta = [-1, 1]$ is equipped with a *density* and that we are allowed to improve the solution discarding a subset $B \subset \Delta$ with small probability $\mathsf{P}[B] = \varepsilon$. The corresponding chance-constrained problem is

$$\begin{aligned} &\min_{(x,y) \in \mathbb{R}^2} \; x + y \\ &\text{subject to} \quad \mathsf{P}\left[\{\delta \in \Delta \;:\; (x - \delta)^2 + y^2 \leq 4\}\right] \geq 1 - \varepsilon. \end{aligned} \tag{4}$$

The setup is shown in Figure 1(b): the feasible set has been enlarged, and the solution has improved a bit, at the cost of neglecting the set $B \subset \Delta$ depicted in light gray. The balls Θ_δ with a center $(\delta, 0)$ such that $\delta \in B$ (partially visible on the right with light gray background) do *not* contain the solution •; the main point of chance-constrained programming is that the probability ε that one such ball pops up in reality, dooming the solution to be wrong, is small; in real-world applications this *risk* is often acceptable. □

Chance-constrained programming is now a well-known subject in stochastic optimization; it has been studied systematically for the first time in the work of Prékopa (see, e.g., the seminal work [14] and the references therein; see [17, Chapter 1] for an introduction and some motivating examples). A well-known drawback of problem (2) is that it is usually hard to solve, because its feasible set is not necessarily convex despite the convexity of the sets Θ_δ. (In this respect, Example 1 is *really* a toy problem.) Another aspect of (2) that seems innocuous but that I consider a drawback—being often unrealistic in practice—is that it requires the exact knowledge of P and of the mapping $\delta \mapsto \Theta_\delta$.

Another way to weaken problem (1) is, in Marco Campi's words, to "let the data speak," to solve a *random problem with finitely many constraints*, and to provide a high-confidence guarantee on its solution. This method is called the *scenario approach*

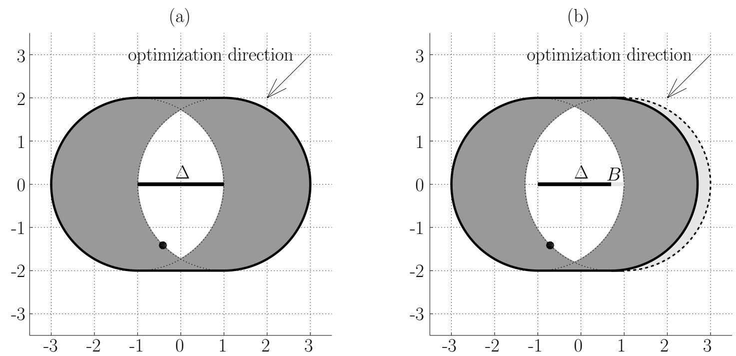

FIG. 1. *Example* 1*:* (a) *Robust problem* (3)*,* (b) *Chance-constrained problem* (4).

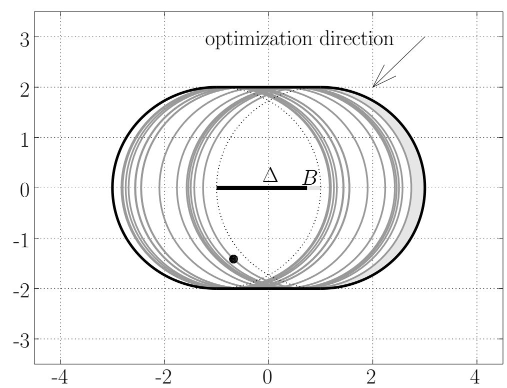

FIG. 2. *Instance of scenario program with sample size $N = 15$; the solution θ^*_N is marked with* •*. The centers $(\delta, 0)$, $\delta \in B$, of the balls Θ_δ that do* not *contain θ^*_N are plotted in light gray.*

and was introduced in [3], initially aiming at robust control design. (See also [7] for a nice introduction.) Its fundamental ideas develop as follows: it is supposed that the experimenter can observe a finite sample of independent and identically distributed constraint sets $\{\Theta_{\delta^{(i)}}\}_{i=1}^N$, extracted according to P^N; s/he forms the problem

$$
\begin{aligned}
&\min_{\theta \in \Theta} \; c^\top \theta \\
&\text{subject to} \quad \theta \in \Theta_{\delta^{(i)}} \text{ for all } i = 1, \ldots, N
\end{aligned} \tag{5}
$$

and computes its solution θ^*_N. Each element $\Theta_{\delta^{(i)}}$ of the finite sample is called a *scenario*, and problem (5) is called a *scenario program*; an instance of (5), along the lines of Example 1, is shown in Figure 2. *Before observing the constraints*, the solution θ^*_N may be regarded as a random vector over Δ^N, and hence the set $B \subset \Delta$ mapping to constraints $\{\Theta_\delta\}_{\delta \in B}$ that do not contain θ^*_N is a random set, and its probability $\mathsf{P}[B]$ is a random variable over Δ^N. The scenario approach attaches to θ^*_N a certificate of this form:

the probability that $\mathsf{P}[B] \leq \varepsilon$ *is always greater than or equal to* $1 - \beta$,

where β is a parameter that depends only on the sample size N and the dimension d of the problem, and that decreases very quickly as the sample size increases. *After* *observing the constraints*, when the solution θ^*_N has been computed, the experimenter can claim that $\mathsf{P}[B] \leq \varepsilon$ with *confidence* $1-\beta$.[1]

The scenario approach has a number of advantages over chance-constrained programming. First, every realization of problem (5) is convex and, the sample of constraint sets being finite, usually simple to solve. Second, for any fixed $\varepsilon \in (0,1)$, the sample size N has a logarithmic dependence on the confidence parameter β ($N \sim \log(1/\beta)$), and hence to decrease β of k *orders of magnitude* it is sufficient to increase N by a *factor* k; this allows one to attain *very* high confidence (e.g., $1-\beta = 1-10^{-10}$) or, so to say, "practical certainty," with a relatively small sample size N. Third, and most important, the true fundamental assumption of the scenario approach is just that $\delta^{(1)}, \ldots, \delta^{(N)}$ are independent and identically distributed; except for this, the guarantees on θ^*_N provided by the various developments of the theory are all *universal*, in the sense that they hold irrespective of P and of the mapping $\delta \mapsto \Theta_\delta$. (In other words, the knowledge of P and $\delta \mapsto \Theta_\delta$ is not required.) Furthermore, the theory of convex scenario optimization provides insight about chance-constrained programming: indeed a "hot" research topic is the connection between the probabilistic guarantees on the solution of (5) and the feasibility of (2); see, e.g., [5], [13] for recent developments. To formalize the above discussion, consider the following definition.

Definition 1. *Let $\theta \in \Theta$. The point θ* violates *the constraint set Θ_δ if $\theta \notin \Theta_\delta$. The* violation probability *of θ is defined as follows:*

$$\mathsf{V}(\theta) = \mathsf{P}[\{\delta \in \Delta \ : \ \theta \notin \Theta_\delta\}].$$

*For any $\varepsilon \in (0,1)$, θ is said to be ε-*robust *if $\mathsf{V}(\theta) \leq \varepsilon$.* □

When this does not generate ambiguity, throughout the paper I will adopt the notation $\mathsf{P}[\theta \in A_\delta]$ and $\mathsf{P}[f_\delta(\theta) \in B]$ to denote the probabilities $\mathsf{P}[\{\delta \in \Delta : \theta \in A_\delta\}]$ and $\mathsf{P}[\{\delta \in \Delta : f_\delta(\theta) \in B\}]$, respectively, it being understood that $\mathsf{P}[\cdot]$ captures the variable δ. With this convention the violation probability of θ reads $\mathsf{V}(\theta) = \mathsf{P}[\theta \notin \Theta_\delta]$.

Consider problem (5), and denote θ^*_N its solution; since θ^*_N is a random variable over $(\Delta^N, \mathcal{F}^N, \mathsf{P}^N)$, so is its violation probability $\mathsf{V}(\theta^*_N)$. (Measurability issues are addressed in [13].) The following theorem is a milestone of the scenario approach with convex constraints.

Theorem 2 (Campi, Garatti). *For any $\varepsilon \in (0,1)$, irrespective of P and of the mapping $\delta \mapsto \Theta_\delta$, the following bound holds:*

$$\mathsf{P}^N[\theta^*_N \text{ exists and } \mathsf{V}(\theta^*_N) > \varepsilon] \leq \sum_{k=0}^{d-1} \binom{N}{k} \varepsilon^k (1-\varepsilon)^{N-k} =: \beta. \tag{6}$$ □

For the proof of Theorem 2, the reader is referred to the paper [4].[2] It is shown, there, that the bound (6) is tight, i.e., that there exists a class of so-called nondegenerate problems for which (6) holds with equality; for such problems, $\mathsf{V}(\theta^*_N)$ is a random variable with Beta$(d, N+1-d)$ density, irrespective of P. (For nondegenerate problems

[1] For comparison, recall that in chance-constrained programming $\mathsf{P}[B] = \varepsilon$ is a constant, fixed in advance.

[2] Strictly speaking, Theorem 2 does not require that the solution θ^*_N to problem (5), when it exists, is *unique*; nevertheless, it is customary in the scenario approach literature to assume that it is possible to isolate a single solution, when there are many, by means of a "tie-break" rule. I will adhere to this convention and refer to "*the* solution θ^*_N" rather than "*a* solution θ^*_N," although in the following this will be needed only to make Theorem 2 work properly.

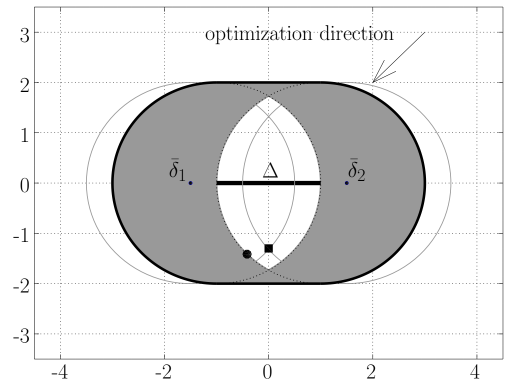

FIG. 3. *Solution of the robust problem (■) and of the* essential *robust one (•).*

in min-max form—see section 2.1—much more is known; see, e.g., [9] for a recent more general result.) The logarithmic dependence between β and N descends from the definition $\beta := \sum_{k=0}^{d-1} \binom{N}{k} \varepsilon^k (1-\varepsilon)^{N-k}$; for a detailed discussion see, e.g., the introduction of [4] and the comparison with [3] therein.

Theorem 2 says that when θ_N^* exists it is ε-robust with confidence $1-\beta$. If ε is small and N big enough so that the confidence $1-\beta$ is very high—"practical certainty"—one can say that the scenario minimum $c^\top \theta_N^*$ is "close enough" to the "robust minimum" (from a risk-analysis perspective, albeit not necessarily in the metric sense); and I dare say, in statistical jargon, that $c^\top \theta_N^*$ is a good *estimator* of the "robust minimum." I claim that, under fairly general hypotheses, this estimator is also *consistent*—hence the title of the paper—i.e., $c^\top \theta_N^*$ converges to the robust minimum as $N \to \infty$. But there is a big *caveat* here. Unless the probability P and the mapping $\delta \mapsto \Theta_\delta$ are particularly well-behaved, the true robust minimum is not the solution of problem (1); it is instead the solution of the following one, that I like to call the *essential robust problem*:

$$\begin{aligned} &\min_{\theta \in \Theta} \ c^\top \theta \\ &\text{subject to} \quad \mathsf{P}\left[\theta \in \Theta_\delta\right] = 1. \end{aligned} \tag{7}$$

The meaning of the solution of (7) and its fundamental difference with the solution of (1) are illustrated in Figure 3, along the lines of Example 1. Suppose that the sample space $\Delta = [-1, 1]$ of Example 1, equipped, e.g., with a uniform density, is augmented with two points $\bar{\delta}_1 = -3/2$ and $\bar{\delta}_2 = 3/2$ in such a way that $\mathsf{P}\left[\{\bar{\delta}_1\}\right] = \mathsf{P}\left[\{\bar{\delta}_2\}\right] = 0$. Despite the fact that the balls centered at $(\bar{\delta}_1, 0)$ and $(\bar{\delta}_2, 0)$ show up with probability 0 and are completely *inessential*, the original robust problem must take them into account; its solution is marked with ■ in the plot. On the other hand, the essential robust problem (7) disregards the negligible event $\{\bar{\delta}_1, \bar{\delta}_2\}$; its solution is marked with •.

To make my claim more rigorous, let θ^{**} be a solution of the essential robust problem (7). By saying that θ_N^* is *consistent* I mean that, under fairly general assumptions, $c^\top \theta_N^* \to c^\top \theta^{**}$ almost surely as $N \to \infty$, and that if θ^{**} is unique (• in Figure 3), then also $\theta_N^* \to \theta^{**}$ almost surely. *This is precisely the main message of the paper.* The message can, and will, be translated in the language of min-max optimization; such a translation will lead to more intuitive assumptions and to simpler proofs; besides, it will make some justice of the otherwise arbitrary adjective "essential."

Structure of the paper. Section 2 is meant to show that the solution of the scenario program (5) is equivalent to the solution of a min-max problem, and to the introduction of a function $\mathbf{L}$ whose minimization is equivalent to the solution of the essential robust problem (7). Section 3 is dedicated to the proof of some fundamental properties of $\mathbf{L}$ (to start with, convexity and lower semicontinuity). Section 4 contains, among other properties of the function $\mathbf{L}$ and of the violation probability V, the rigorous statement and the proof of the main results of the paper (Theorems 14 and 15). In section 5 I "translate" the assumptions of Theorems 14 and 15, which are stated with respect to min-max optimization, back to the language of problems (5)–(7), and restate the main result with respect to these problems (Theorem 17). The discussion preceding and following Theorem 17 shows that $c^\top \theta_N^* \to +\infty$ if and only if problem (7) is infeasible, and that in turn this happens if and only if $\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} = \varnothing$ almost surely. Section 6 concludes the paper with some final remarks and acknowledgments.

2. Formalization of the problem: min-max optimization. To establish the main convergence results, I find it convenient to express constraints in terms of convex functions rather than convex sets, and to recast scenario programs as min-max programs. In this section I will show that the two approaches are equivalent, then show that the concept of max needed to recast the robust problem is somewhat subtler than one would expect at first sight, formalize two assumptions that will be used throughout the paper, state the main results, and try to provide some insight about the assumptions and their consequences, along with a brief comparison with some recent results in the literature.

2.1. Scenario min-max problems. Let $(\Delta, \mathcal{F}, \mathsf{P})$ and $(\Delta^N, \mathcal{F}^N, \mathsf{P}^N)$ be the probability spaces defined at the beginning of section 1, let $\mathcal{X}$ be a closed convex subset of $\mathbb{R}^d$, and suppose that to each $\delta \in \Delta$ there is associated a convex function $f_\delta : \mathcal{X} \to \overline{\mathbb{R}}$, taking values in the *extended* real set $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. Consider the following min-max problem:

$$
\begin{aligned}
&\text{let} \quad \hat{f}_N(x) = \max_{i=1\cdots N} f_{\delta^{(i)}}(x); \\
&\text{find} \quad y_N^* = \min_{x\in\mathcal{X}} \hat{f}_N(x), \quad x_N^* = \arg\min_{x\in\mathcal{X}} \hat{f}_N(x),
\end{aligned} \tag{8}
$$

where $(\delta^{(1)}, \dots, \delta^{(N)}) \in \Delta^N$.[3] The epigraphical form of (8) is as follows:

$$
\begin{aligned}
&\min_{(x,y)\in(\mathcal{X}\times\mathbb{R})} \quad y \\
&\text{subject to} \quad f_{\delta^{(i)}}(x) \le y \quad \text{for all } i = 1, \dots, N.
\end{aligned} \tag{9}
$$

On one hand, problem (9) is a particular, $(d+1)$-dimensional instance of problem (5); this follows immediately letting $\Theta = \mathcal{X} \times \mathbb{R}$, $\theta = (x, y)$, $c^\top \theta = y$, and $\Theta_\delta = \{(x, y) \in \mathcal{X} \times \mathbb{R} : f_\delta(x) \le y\}$, and assuming, by convention, that $+\infty$ must be understood as the minimum of problem (9) if such problem is infeasible. On the other hand, every problem in the form (5) can be formulated as a min-max problem in the form (8). To

[3]Following the literature of the scenario approach, I assume that a single solution $x_N^* \in \mathcal{X}$ can be isolated by means of a "tie-break" rule even if many solutions exist, although this is not really relevant to my discussion and is only necessary to make Theorem 2 work properly. Thus, in the rest of the paper $\arg\min_{x\in\mathcal{X}} \hat{f}_N(x)$ will always denote an *element* of $\mathcal{X}$, not a *subset* of $\mathcal{X}$ as it would naturally mean.

do this, let $\mathcal{X} = \Theta$, $x = \theta$, and for all $\delta \in \Delta$ define $f_\delta : \Theta \to \overline{\mathbb{R}}$ as follows:[4]

$$f_\delta(\theta) = c^\top \theta + \mathcal{I}_{\Theta_\delta}(\theta) = \begin{cases} c^\top \theta & \text{if } \theta \in \Theta_\delta, \\ +\infty & \text{otherwise.} \end{cases}$$

Then problems (5), (8), and (9) are equivalent, i.e., they yield exactly the same solutions $x_N^* = \theta_N^*$, $y_N^* = c^\top \theta_N^*$, and hence any result about the convergence of the solution (x_N^*, y_N^*) of (8), as $N \to \infty$, reflects on a similar result about the convergence of the solution $(\theta_N^*, c^\top \theta_N^*)$ of (5), and vice versa.

In the setting of problem (9), the violation probability of (x, y) reads $\mathsf{V}(x, y) = \mathsf{P}\left[f_\delta(x) > y\right]$, and since problem (9) is $(d+1)$-dimensional, the bound established by Theorem 2, provided that the solution (x_N^*, y_N^*) exists, becomes

$$\mathsf{P}^N\left[\mathsf{V}(x_N^*, y_N^*) > \varepsilon\right] \le \sum_{k=0}^{d} \binom{N}{k} \varepsilon^k (1-\varepsilon)^{N-k}. \tag{10}$$

Since Theorem 2 lays the foundation for the main results of this paper, and since for the bound (10) to make sense it is necessary to ensure that a solution of (8) exists at least for N big enough, I need to introduce two assumptions that will be sufficient for this to hold. The first assumption goes as follows.[5]

Assumption 1. *The domain $\mathcal{X} \subseteq \mathbb{R}^d$ is convex and closed; for all $\delta \in \Delta$, $f_\delta : \mathcal{X} \to \overline{\mathbb{R}}$ is convex and lower semicontinuous. For all $x \in \mathcal{X}$, $\delta \mapsto f_\delta(x)$ is P-measurable (i.e., $f_{(\cdot)}(x)$ is a random variable).* □

An immediate consequence of Assumption 1 is that $\hat{f}_N$ is convex and lower semicontinuous for all $N \in \mathbb{N}$. Without further mention, let us agree that Assumption 1 will be in force throughout the whole paper. Here follows the second assumption.

Assumption 2. *For all $N \in \mathbb{N}$, $\hat{f}_N$ is P^N-almost surely proper.*[6] *Moreover, there exists $\bar{N} \in \mathbb{N}$ such that*

$$\mathsf{P}^{\bar{N}}\left[\hat{f}_{\bar{N}} \text{ is coercive}\right] > 0. \qquad \square$$

Remark. If the functions f_δ are themselves coercive for all $\delta \in \Delta$, then the coercivity of $\hat{f}_N$ follows automatically for all $N \in \mathbb{N}$. There are conditions that ensure this property, which are very easy to check but otherwise rather conservative. One such condition is, of course, that the domain $\mathcal{X}$ is compact. Another condition, assuming

[4] Here $\mathcal{I}_A(x)$ is the "indicator function of convex analysis," which takes the value 0 if $x \in A$ and the value $+\infty$ otherwise. Later in the paper the "indicator function of probability theory" $\mathbb{1}_A(x)$, taking the value 1 if $x \in A$ and the value 0 otherwise, will also show up and be denoted $\mathbb{1}(x \in A)$ for the sake of readability.

[5] For future reference, let me recall here some standard terminology and notation about $\overline{\mathbb{R}}$-valued functions. A function $F : \mathcal{X} \to \overline{\mathbb{R}}$ is *lower semicontinuous* if, for all $\bar{x} \in \mathcal{X}$, $F(\bar{x}) \le \liminf_{x \to \bar{x}} F(x)$. For $t \in \mathbb{R}$, the *t-sublevel set* of F is the set $\{x \in \mathcal{X} : F(x) \le t\}$. The following are well-known facts (see, e.g., [12]): F is lower semicontinuous if and only if all its t-sublevel sets are *closed*; the pointwise supremum $F(x) = \sup_{\alpha \in A} F_\alpha(x)$ of an arbitrary family of lower semicontinuous functions F_α is lower semicontinuous. The *effective domain* of a function $F : \mathcal{X} \to \overline{\mathbb{R}}$ is the set $\operatorname{dom} F = \{x \in \mathcal{X} : F(x) < +\infty\}$; the *closure* of $\operatorname{dom} F$ is denoted $\overline{\operatorname{dom}}\, F$. A function $F : \mathcal{X} \to \overline{\mathbb{R}}$ is *proper* if it has a nonempty effective domain, i.e., if it does not take the only value $+\infty$. A function $F : \mathcal{X} \to \overline{\mathbb{R}}$ is *coercive* if for all $t \in \mathbb{R}$ there exists a compact set $C \subseteq \mathcal{X}$ including the t-sublevel set of F: $\{x \in \mathcal{X} : F(x) \le t\} \subseteq C$. Of course, if $\mathcal{X}$ is compact, then $F : \mathcal{X} \to \overline{\mathbb{R}}$ is automatically coercive.

[6] The assumption that $\hat{f}_N$ is almost surely proper implies that f_δ is proper for almost all $\delta \in \Delta$. In general the converse is false: for instance, if $f_\delta(x) = \mathcal{I}_{[\delta, \delta+1]}(x)$, where δ takes the values 0 and 2 with equal probability 1/2, then $\hat{f}_2(x) = +\infty$ for all $x \in \mathbb{R}$ with probability 1/2.

for simplicity that $\mathcal{X} = \mathbb{R}^d$, is that $\lim_{\|x\|\to\infty} f_\delta(x) = +\infty$ for all $\delta \in \Delta$.[7] But here we need substantially less: a practical way to check Assumption 2 is indeed to isolate finitely many events, say, $A_1, \dots, A_{\bar{N}} \subseteq \Delta$, each with positive probability, such that for any choice of $\delta^{(i)} \in A_i$, $i = 1, \dots, \bar{N}$, the function $\hat{f}_{\bar{N}}(x) = \max_{i=1,\dots,\bar{N}} f_{\delta^{(i)}}(x)$ tends to $+\infty$ (or "bumps" against the boundary of $\mathcal{X}$) as $\|x\| \to \infty$. $\square$

Example 2. Let $\delta = (a, b)$, where a and b are both random variables with uniform density in $[-1, 1]$. Let $\mathcal{X} = \mathbb{R}$ and $f_\delta : \mathbb{R} \to \mathbb{R}$ be defined as follows:

$$f_\delta(x) = ax + b.$$

Clearly, $\hat{f}_1(x) = f_{\delta^{(1)}}(x)$ (an affine function) cannot be coercive, since all its sublevel sets are unbounded (or empty, in the negligible event $a^{(1)} = 0$). On the other hand, $\hat{f}_2(x) = \max\{f_{\delta^{(1)}}(x), f_{\delta^{(2)}}(x)\}$ is coercive when $f_{\delta^{(1)}}$ and $f_{\delta^{(2)}}$ have opposite slopes (then $\hat{f}_2$ is a "V-shaped" function). This happens with probability $1/2$, and hence Assumption 2 holds with $\bar{N} = 2$. In view of the above remark, the key property here is that the functions f_δ with positive slope (event A_1, with probability $1/2$) tend to $+\infty$ when $x \to +\infty$, and those with negative slope (event A_2, also with probability $1/2$) tend to $+\infty$ when $x \to -\infty$; these are the only directions along which x can tend to infinity when $\mathcal{X} = \mathbb{R}$. $\square$

2.2. A "meaningful min-max": Main results of the paper. The objective of this paper is to prove the almost sure convergence of the empirical minimum y_N^* to, roughly speaking, the "min-max of all the functions f_δ for $\delta \in \Delta$." But what is "the maximum of all the functions f_δ," exactly, supposed to mean, since it is clear that a true maximum may not even exist?[8] The first rigorous answer that comes to mind is, of course,

$$\mathbf{S}(x) = \sup_{\delta\in\Delta} f_\delta(x).$$

$\mathbf{S}$ has many nice properties that one would demand from stochastic optimization: it is convex, lower semicontinuous, and by construction $\hat{f}_N(x) \leq \mathbf{S}(x)$ for all $x \in \mathcal{X}$. Since the $\hat{f}_N$'s form a nondecreasing sequence of functions, it is licit to wonder whether, as $N \to \infty$, $\hat{f}_N \to \mathbf{S}$ in some sense, and maybe whether

$$y_N^* = \min_{x\in\mathcal{X}} \hat{f}_N(x) \to \min_{x\in\mathcal{X}} \mathbf{S}(x). \tag{11}$$

Unfortunately, in any useful probabilistic sense the claim (11) is in general *false*, as the following example shows.

Example 3. Suppose that δ is a random variable with uniform density in $[0, 1]$. Let $\mathcal{X} = \mathbb{R}$ and $f_\delta : \mathbb{R} \to \mathbb{R}$ be defined by

$$f_\delta(x) = \begin{cases} |x| + 2 - \delta & \text{if } \delta = 1/n \text{ for some } n \in \mathbb{N}, \\ |x| + \delta & \text{otherwise.} \end{cases} \tag{12}$$

Here $\mathbf{S}(x) = |x|+2$ and $\min_{x\in\mathcal{X}} \mathbf{S}(x) = 2$. However, since δ has a density, it holds that $\mathsf{P}[\delta = 1/n \text{ for some } n \in \mathbb{N}] = 0$, and hence $\hat{f}_N(x) > |x| + 1$ happens with probability 0, and almost surely $\min_{x\in\mathcal{X}} \hat{f}_N(x) \leq 1$ for all $N \in \mathbb{N}$. A pictorial view of the family $\{f_\delta\}_{\delta\in\Delta}$ and of $\mathbf{S}$ is shown in Figure 4. $\square$

---

[7]Indeed, assume that $\lim_{\|x\|\to\infty} f_\delta(x) = +\infty$. Then, by definition of limit, for all $t \in \mathbb{R}$ there exists $M > 0$ such that $f_\delta(x) > t$ when $\|x\| > M$. Hence the sublevel set $\{x \in \mathcal{X} : f_\delta(x) \leq t\}$ is a subset of the closed ball $\{x \in \mathcal{X} : \|x\| \leq M\}$, which is always compact since $\mathbb{R}^d$ is finite-dimensional.

[8]In the following Example 3, for instance, $\max_{\delta\in\Delta} f_\delta(x)$ does not exist for any $x \in \mathcal{X}$.

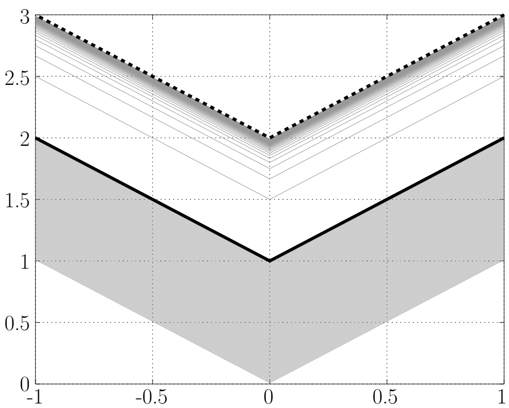

FIG. 4. *Gray lines: functions $f_\delta(x)$ defined in* (12)*; dashed black line: supremum $\mathbf{S}(x)$; solid black line: essential supremum $\mathbf{L}(x)$ (see Definition* 3*).*

Since in the scenario approach P and the functions f_δ (or the sets Θ_δ) are unknown, we cannot exclude that even a single $f_{\bar{\delta}}$, extracted with probability 0, dominates all the other functions f_δ, $\delta \in \Delta$, thus driving the minimum of $\mathbf{S}$ by itself alone. Indeed $\mathbf{S}$ is a "useless supremum." The useful supremum is instead the upper boundary of the region where all the mass of the functions lies. In Example 3, all the mass of the functions lies below the upper bound $|x| + 1$, which is actually the *least* such upper bound; anything above $|x| + 1$ is negligible and must be discarded in our discussion. The reader can now safely expect that the minimum of $\hat{f}_N$ converges to $\min_{x\in\mathbb{R}} |x| + 1 = 1$. In fact this is what happens almost surely: $|x| + 1$ is the "useful supremum"! This example leads to the following definition.

DEFINITION 3. *Let the function $\mathbf{L} : \mathcal{X} \to \overline{\mathbb{R}}$ be defined as follows:*

$$\mathbf{L}(x) = \inf \{y \in \mathbb{R} \ : \ \mathsf{P}\,[f_\delta(x) > y] = 0\}\,.$$

In measure theory $\mathbf{L}(x)$ is known as the *essential supremum* of the function $\delta \mapsto f_\delta(x)$ and denoted $\operatorname{ess\,sup}_{\delta\in\Delta} f_\delta(x)$. Thus, $\mathbf{L}$ is the *(x-)pointwise essential supremum* of the family of functions $\{f_\delta\}_{\delta\in\Delta}$. The function $\mathbf{L}$ is precisely the useful supremum: it is immediate to check that, in Example 3, $\mathbf{L}(x) = |x| + 1$ (see Figure 4, solid black line). Note that, even if $\hat{f}_N$ is proper for all N and for any choice of $(\delta^{(1)}, \dots, \delta^{(N)})$, $\mathbf{L}$ (as well as $\mathbf{S}$) can take the value $+\infty$ for all $x \in \mathcal{X}$. The main results of the paper (Theorems 14 and 15) assert that, under Assumptions 1 and 2,

- if $\mathbf{L} \equiv +\infty$, then $y^*_N \to +\infty$ almost surely;[9]
- otherwise $\min \mathbf{L} < +\infty$ exists, and $y^*_N \to \min \mathbf{L}$ almost surely;
- if, moreover, $\arg\min \mathbf{L}$ is unique, then $x^*_N \to \arg\min \mathbf{L}$ almost surely.

For the same reasons why, in general, the solution y^*_N does not converge to $\min \mathbf{S}$, in general the minimum $c^\top \theta^*_N$ attained by the solution of (5) does not converge to the minimum of problem (1). It converges instead to the minimum attained by the solution—now let me call it *essential robust solution*—of problem (7). The connection with the main results is fairly intuitive; all it requires is to translate the meaning of $\mathbf{L}$, and Assumptions 1 and 2, to the domain $\theta, \Theta, \Theta_\delta$ of problems (5)–(7). Once this job

[9] More rigorously, P^∞-almost surely, where P^∞ is a probability on the space of infinite sequences of elements in Δ, compatible with P^N for all N; for more details, refer to the beginning of section 4.

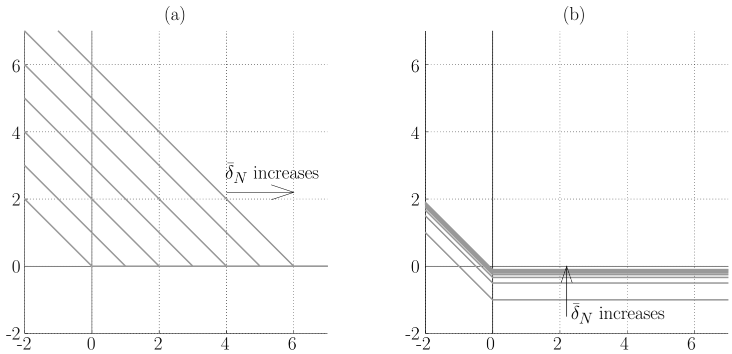

FIG. 5. *Example* 4: (a) *functions* $f_\delta(x)$ *as in* (13), (b) *functions* $f_\delta(x)$ *as in* (14).

is done, the main results of the paper can be stated as follows (Theorem 17): under the equivalent of Assumptions 1 and 2,

- if problem (7) is infeasible, then $c^\top \theta_N^* \to +\infty$ almost surely;
- otherwise, a solution θ^{**} of (7) exists and $c^\top \theta_N^* \to c^\top \theta^{**}$ almost surely;
- if, moreover, θ^{**} is unique, then $\theta_N^* \to \theta^{**}$ almost surely.

I will discuss in detail this version of the main results in section 5.

2.3. Discussion of the assumptions and comparison with literature. Suppose that Assumption 1 holds and that $\hat{f}_N$ is proper and coercive for N big enough. Since $\hat{f}_N$ is proper, coercive, and lower semicontinuous, it attains a finite minimum in $\mathcal{X}$. This is the assertion of the Tonelli–Weierstrass theorem (see, e.g., [2, Proposition 3.2.1]), a generalization of the "classical" Weierstrass's theorem. But actually the true targets of Assumption 2 are the coercivity of $\mathbf{L}$ (Proposition 11) and the existence of compact sublevel sets. From my point of view, Assumption 1 is natural in any convex optimization problem; but Assumption 2 is truly the substantial requirement here, and I conjecture that *nothing* can be established about the convergence of y_N^* if the $\hat{f}_N$'s are not, in a way or another, coercive for N big enough, and if $\mathbf{L}$ does not have compact sublevel sets. The following example illustrates the kind of issues that may arise.

Example 4. Let $p \in (0,1)$ and δ be a random variable with geometric distribution $\mathsf{P}[\delta = n] = p(1-p)^n$. Let $\mathcal{X} = \mathbb{R}$ and $f_\delta : \mathbb{R} \to \mathbb{R}$ be defined by

$$f_\delta(x) = \begin{cases} \delta - x & \text{if } x \le \delta, \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

(see Figure 5(a)). Letting $\bar{\delta}_N = \max_{i=1,\dots,N} \delta^{(i)}$, we have $\hat{f}_N(x) = \bar{\delta}_N - x$ for $x \le \bar{\delta}_N$, 0 otherwise. No function $\hat{f}_N$ is coercive, and convergence fails: while $y_N^* = \min_{x \in \mathcal{X}} \hat{f}_N(x) = 0$ for all N and for any possible extraction of $\delta^{(1)}, \dots, \delta^{(N)}$, it is immediate to recognize that $\mathbf{L}(x) = +\infty$ for all $x \in \mathcal{X}$. Let instead $f_\delta : \mathbb{R} \to \mathbb{R}$ be defined by

$$f_\delta(x) = \begin{cases} -1/(\delta+1) - x & \text{if } x \le 0, \\ -1/(\delta+1) & \text{otherwise} \end{cases} \tag{14}$$

(see Figure 5(b)). In this case, letting $\bar{\delta}_N = \max_{i=1,\dots,N} \delta^{(i)}$ as before, it holds that $\hat{f}_N(x) = -1/(\bar{\delta}_N + 1) - x$ for $x \le 0$, $-1/(\bar{\delta}_N + 1)$ otherwise, and $\mathbf{L}(x) = -x$ for

$x \leq 0$, 0 otherwise. Neither $\mathbf{L}$ nor any $\hat{f}_N$ is coercive, but this time $y_N^* = -1/(\bar{\delta}_N + 1)$ converges to $0 = \min \mathbf{L}$ almost surely. $\square$

On the same subject, it is worth comparing the results presented here with the work of Shapiro and others on sample average approximation (SAA) estimators; see, e.g., [11], [17]. Roughly speaking, this paper proves for the min-max problem what in [17, section 5.1.1] is proved about the minimum of the sample average (of random real-valued convex functions). Letting $\bar{f}_N(x) = \frac{1}{N}\sum_{i=1}^N f_{\delta^{(i)}}(x)$ and $\mathbf{F}(x) = \mathsf{E}\left[f_\delta(x)\right]$ it is shown there that under fairly general hypotheses, as $N \to \infty$, $\min \bar{f}_N \to \min \mathbf{F}$ and $\arg\min \bar{f}_N \to \arg\min \mathbf{F}$ almost surely. One of these hypotheses recurs in different forms: [17, Theorem 5.3], "there exists a *compact* set $C \in \mathbb{R}^d$ such that [...] the set S of optimal solutions of the true problem is nonempty and is contained in C," or [17, Theorem 5.4], "the set S of optimal solutions of the true problem is nonempty and *bounded*." (The "true problem" is to find [arg] min $\mathbf{F}$.) *Both these requirements go in the same direction as the coercivity of* $\mathbf{F}$: we all need the existence of compact sublevel sets!

However, differently from SAA, neither the law of large numbers nor the central limit theorem applies to min-max stochastic problems; hence, whereas averaging typically ensures that the mean square error (MSE, that is; the expected squared deviation from the "true" minimum) decreases with rate $1/N$ as happens for SAA [17, Theorem 5.7], here the MSE's decrease rate can be arbitrarily low; in other words, although $y_N^* = \min \hat{f}_N$ does converge to $\min \mathbf{L}$ almost surely as $N \to \infty$, the convergence can be arbitrarily slow, as the following example shows.

Example 5. Let $a > 0$ and $b = 1/a$. Suppose that δ is a random variable with exponential density $g_\delta(t) = ae^{-at}$ for $t \in [0, +\infty)$. Let $\mathcal{X} = [-1, 1]$ and $f_\delta : \mathcal{X} \to \mathbb{R}$ be defined by

$$f_\delta(x) = x^2 - e^{-\delta/2}.$$

Here $\mathbf{L}(x) = x^2$ and $\min_{x\in\mathcal{X}} \mathbf{L}(x) = 0$; letting $\bar{\delta}_N = \max_{i=1,\dots,N} \delta^{(i)}$, it holds that $\hat{f}_N(x) = x^2 - e^{-\bar{\delta}_N/2}$ and $y_N^* = \min_{x\in\mathcal{X}} \hat{f}_N(x) = -e^{-\bar{\delta}_N/2}$. The cumulative distribution function of each $\delta^{(i)}$ is $G_{\delta^{(i)}}(t) = \mathsf{P}\left[\delta^{(i)} \leq t\right] = 1 - e^{-at}$, and hence the cumulative distribution function of $\bar{\delta}_N$ is $G_{\bar{\delta}_N}(t) = \mathsf{P}^N\left[\bar{\delta}_N \leq t\right] = (1 - e^{-at})^N$, and its density is $g_{\bar{\delta}_N}(t) = Na(1 - e^{-at})^{N-1}e^{-at}$. The MSE of y_N^* with respect to its own limit is

$$\begin{aligned}
&\mathsf{E}\left[(\min \mathbf{L} - y_N^*)^2\right] = \mathsf{E}\left[e^{-\bar{\delta}_N}\right]\\
&\quad= \int_0^{+\infty} e^{-t}\ Na(1-e^{-at})^{N-1}e^{-at}\ dt \quad (\text{let } u = e^{-at})\\
&\quad= N\int_0^1 u^{1/a}(1-u)^{N-1}\ du \quad (= N\times \text{ Euler's Beta}(1/a+1, N) \text{ function})\\
&\quad= \frac{\Gamma(b+1)\ \Gamma(N+1)}{\Gamma(N+b+1)} \quad (\text{now use Stirling's approximation})\\
&\quad\sim \frac{\sqrt{2\pi N}\ N^N\ e^{-N}}{\sqrt{2\pi(N+b)}\ (N+b)^{(N+b)}\ e^{-(N+b)}}\ \Gamma(b+1)\\
&\quad= \sqrt{\frac{N}{N+b}}\cdot\left(\frac{N}{N+b}\right)^N\cdot\frac{1}{(N+b)^b}\cdot e^b\cdot\Gamma(b+1)\\
&\quad\sim 1\cdot\frac{1}{e^b}\cdot\frac{1}{(N+b)^b}\cdot e^b\cdot\Gamma(b+1) \sim \frac{\Gamma(b+1)}{N^b}.
\end{aligned}$$

Since the parameter $b = 1/a$ can be chosen arbitrarily small, the decrease rate MSE $\sim \frac{1}{N^b}$ can be arbitrarily low. $\square$

The search for *performance bounds* on the approximation error with respect to the robust solution, which in min-max form would read $\min \mathbf{S} - y_N^*$, is another "hot" research topic in the scenario approach literature. Letting aside the question of what are the minimal requirements on P and on the mapping $\delta \mapsto f_\delta$ sufficient to ensure that $\mathbf{L} = \mathbf{S}$, which may be interesting *per se*,[10] I would like to mention that results of this kind, coming in the form

$$\text{for big enough } N, \quad \mathsf{P}^N\left[\min_{x\in\mathcal{X}} \mathbf{S}(x) - y_N^* \leq \varepsilon\right] \geq 1-\beta \tag{15}$$

with very high confidence $1-\beta$, are now available; see, for example, the recent paper [13] by Mohajerin Esfahani, Sutter, and Lygeros, partially building on the previous work [10] by Kanamori and Takeda. But everything comes at a price, and the current price for the performance bound (15) is the requirement of significant knowledge about P and $\delta \mapsto f_\delta$. For example, both [13] and [10] assume the *uniform Lipschitz continuity* of $\delta \mapsto f_\delta(x)$ over $\mathcal{X}$[11] In this paper I prefer to assume the least possible knowledge about P and $\delta \mapsto f_\delta$: therefore the distinction between $\mathbf{S}$ and $\mathbf{L}$ must remain, and with only Assumptions 1 and 2 I maintain that there is not and there cannot be any guarantee on the convergence speed of $y_N^* \to \min \mathbf{L}$. With respect to Example 5,

$$\begin{aligned}
&\mathsf{P}^N\left[\min_{x\in\mathcal{X}} \mathbf{L}(x) - y_N^* \geq \varepsilon\right] = \mathsf{P}^N\left[e^{-\bar{\delta}_N/2} \geq \varepsilon\right]\\
&= \mathsf{P}^N\left[\bar{\delta}_N \leq -\log\varepsilon^2\right] = G_{\bar{\delta}_N}(-\log\varepsilon^2) = \left(1-\varepsilon^{2a}\right)^N,
\end{aligned}$$

which may be arbitrarily close to 1 because a may be arbitrarily large; hence no performance bound in the form (15) can be established at all.[12]

3. Fundamental properties of L. This section is dedicated to the proof of some fundamental properties of $\mathbf{L}$ that would be trivial about $\mathbf{S}$, namely, that $\mathbf{L}$ is convex and lower semicontinuous and that $\hat{f}_N \leq \mathbf{L}$ almost surely. The proof of the latter statement requires two technical lemmas about proper, convex, and lower semicontinuous functions that I did not find in the "standard" literature. Let me start the discussion by resuming in a lemma the most intuitive facts about $\mathbf{L}$.

LEMMA 4. *For all $x \in \mathcal{X}$,*
1. $\mathsf{P}[f_\delta(x) > \mathbf{L}(x)] = \mathsf{V}(x, \mathbf{L}(x)) = 0$*;*
2. $\mathsf{P}[f_\delta(x) \leq \mathbf{L}(x)] = 1$*;*
3. *if $\bar{y} < \mathbf{L}(x)$, then $\mathsf{P}[\bar{y} < f_\delta(x) \leq \mathbf{L}(x)] > 0$;*
4. *if $D \subseteq \Delta$ and $\mathsf{P}[D] = 1$, then $\mathbf{L}(x) \leq \sup_{\delta\in D} f_\delta(x)$.*

[10]Let, e.g., (Δ, m) be a metric space, T be the topology induced by m, $\mathcal{F}$ be the Borel σ-algebra generated by T, P be a probability on $(\Delta, \mathcal{F})$, and $f_{(\cdot)}(x) : \Delta \to \mathbb{R}$ be a $(\mathcal{F}, \mathrm{Borel}(\mathbb{R}))$-measurable function. Are there "nice" sufficient conditions such that $\operatorname{ess\,sup}_{\delta\in\Delta} f_\delta(x) = \sup_{\delta\in\Delta} f_\delta(x)$? A simple sufficient condition is that $\delta \mapsto f_\delta(x)$, understood as a function between metric spaces, is *continuous*, and that the support of P is the whole Δ (i.e., *every open subset of Δ has positive probability*); but I guess that this condition can be weakened. Moreover, one could argue that the condition actually needed in the minimization endeavor is that $\mathbf{L}(x) = \mathbf{S}(x)$ at the minimum points x of $\mathbf{L}$: if this holds, since $\mathbf{L}(x) \leq \mathbf{S}(x)$ for all $x \in \mathcal{X}$ anyway, a minimum point of $\mathbf{L}$ is also a minimum point of $\mathbf{S}$.

[11]The paper [13] also ensures the coercivity of all the $\hat{f}_N$'s by assuming that0 $\mathcal{X}$ is compact.

[12]The fundamental requirement of [13] that fails to hold in Example 5, thus preventing us from establishing performance bounds, is that P must have bounded support. In Example 5 the support of P is $[0, +\infty)$.

Proof. To prove point 1 note that, by the definition of $\mathbf{L}$, for all $y > \mathbf{L}(x)$ it holds that $\mathsf{P}[f_\delta(x) > y] = 0$. If $f_\delta(x) > \mathbf{L}(x)$, then there exists $n \in \mathbb{N}$ such that $f_\delta(x) > \mathbf{L}(x) + 1/n$; hence

$$\{\delta \in \Delta \ : \ f_\delta(x) > \mathbf{L}(x)\} \subseteq \bigcup_{n=1}^{\infty} \left\{\delta \in \Delta \ : \ f_\delta(x) > \mathbf{L}(x) + \frac{1}{n}\right\}$$

and therefore

$$\begin{aligned}\mathsf{P}[f_\delta(x) > \mathbf{L}(x)] &\le \mathsf{P}\left[\bigcup_{n=1}^{\infty} \left\{f_\delta(x) > \mathbf{L}(x) + \frac{1}{n}\right\}\right] \\ &\le \sum_{n=1}^{\infty} \mathsf{P}\left[f_\delta(x) > \mathbf{L}(x) + \frac{1}{n}\right] = 0.\end{aligned}$$

Point 2 is now trivial, since the event $\{f_\delta(x) \le \mathbf{L}(x)\}$ is the complement of the event $\{f_\delta(x) > \mathbf{L}(x)\}$. To prove point 3 suppose, for the sake of contradiction, that $\mathsf{P}[\bar{y} < f_\delta(x) \le \mathbf{L}(x)] = 0$. Then

$$\begin{aligned}\mathsf{P}[f_\delta(x) > \bar{y}] &= \mathsf{P}[\{\bar{y} < f_\delta(x) \le \mathbf{L}(x)\} \cup \{f_\delta(x) > \mathbf{L}(x)\}] \\ &= \mathsf{P}[\bar{y} < f_\delta(x) \le \mathbf{L}(x)] + \mathsf{P}[f_\delta(x) > \mathbf{L}(x)] = 0,\end{aligned}$$

and hence $\bar{y} \in \{y \in \mathbb{R} \ : \ \mathsf{P}[f_\delta(x) > y] = 0\}$; but now $\bar{y} < \mathbf{L}(x)$ and the definition of $\mathbf{L}(x)$ yield a contradiction. To prove point 4, let $\bar{y} = \sup_{\delta \in D} f_\delta(x)$. Any $\delta \in \Delta$ such that $f_\delta(x) > \bar{y}$ belongs to the complement of D, and hence $\mathsf{P}[f_\delta(x) > \bar{y}] = 0$, and $\bar{y} \in \{y \in \mathbb{R} : \mathsf{P}[f_\delta(x) > y] = 0\}$. The claim follows from the definition of $\mathbf{L}(x)$. ∎

The following Propositions 5 and 6 establish the most important properties of $\mathbf{L}$, its convexity and its lower semicontinuity. These are the counterparts, for the pointwise *essential* supremum, of two well-known facts: the pointwise *supremum* of an arbitrary family of convex (resp., lower semicontinuous) functions is itself convex (resp., lower semicontinuous).

PROPOSITION 5. $\mathbf{L}$ *is convex.*

Proof. For the sake of contradiction, suppose that $\mathbf{L}$ is not convex, so that there exist $x_1, x_2 \in \mathcal{X}$ and $\lambda \in (0,1)$ such that $\lambda\mathbf{L}(x_1) + (1-\lambda)\mathbf{L}(x_2) < \mathbf{L}(\lambda x_1 + (1-\lambda)x_2)$. By Lemma 4 (point 2) there exist sets $D_1 \subseteq \Delta$, $D_2 \subseteq \Delta$, both with probability 1, such that $f_\delta(x_1) \le \mathbf{L}(x_1)$ for all $\delta \in D_1$ and $f_\delta(x_2) \le \mathbf{L}(x_2)$ for all $\delta \in D_2$; on the other hand, by Lemma 4 (point 3) there exists a set $B \subset \Delta$ with $\mathsf{P}[B] > 0$ such that for all $\delta \in B$

$$\lambda\mathbf{L}(x_1) + (1-\lambda)\mathbf{L}(x_2) < f_\delta(\lambda x_1 + (1-\lambda)x_2) \le \mathbf{L}(\lambda x_1 + (1-\lambda)x_2). \tag{16}$$

Let $\bar{B} = B \cap D_1 \cap D_2$ and note that, since $\mathsf{P}\left[\bar{B}\right] = \mathsf{P}[B] > 0$, $\bar{B}$ is not empty. But for any $\delta \in \bar{B}$ it holds that $f_\delta(x_1) \le \mathbf{L}(x_1)$, $f_\delta(x_2) \le \mathbf{L}(x_2)$, and by convexity of f_δ

$$\begin{aligned}f_\delta(\lambda x_1 + (1-\lambda)x_2) &\le \lambda f_\delta(x_1) + (1-\lambda)f_\delta(x_2) \\ &\le \lambda\mathbf{L}(x_1) + (1-\lambda)\mathbf{L}(x_2),\end{aligned}$$

which is in contradiction with (16). The contradiction stems from the assumption that $\mathbf{L}$ is not convex, and this concludes the proof. ∎

PROPOSITION 6. **L** *is lower semicontinuous.*

Proof. Fix $\bar{x} \in \mathcal{X}$. Let $(x_n)_{n=1}^{\infty}$ be a sequence of points in $\mathcal{X}$ converging to $\bar{x}$; let moreover

$$D_n = \{\delta \in \Delta \ : \ f_\delta(x_n) \le \mathbf{L}(x_n)\}; \qquad \bar{D} = \bigcap_{n=1}^{\infty} D_n.$$

Since $\mathsf{P}[D_n] = 1$ for all n (Lemma 4, point 2), also $\mathsf{P}\left[\bar{D}\right] = 1$. For all $\delta \in \bar{D}$, the lower semicontinuity of f_δ implies

$$f_\delta(\bar{x}) \le \liminf_{n\to\infty} f_\delta(x_n) \le \liminf_{n\to\infty} \mathbf{L}(x_n),$$

and therefore, by Lemma 4 (point 4),

$$\mathbf{L}(\bar{x}) \le \sup_{\delta \in \bar{D}} f_\delta(\bar{x}) \le \liminf_{n\to\infty} \mathbf{L}(x_n).$$

This holds for all the sequences $(x_n)_{n=1}^{\infty}$ converging to $\bar{x}$, and hence **L** is lower semicontinuous at $\bar{x}$; the claim follows since $\bar{x}$ was chosen arbitrarily. □

LEMMA 7. *Suppose that a function* $L : \mathbb{R}^d \to \overline{\mathbb{R}}$ *is proper, convex, and lower semicontinuous. Then for all* $\bar{x} \in \overline{\operatorname{dom} L}$

$$\liminf_{\substack{x\to\bar{x}\\ x\in\operatorname{dom} L}} L(x) = L(\bar{x}).$$

Remark. It is a well-known fact that any proper convex function L is actually *continuous* in the interior of its effective domain [12, Corollary 2.1.3], but here I am particularly interested in what happens at the *boundary* of $\operatorname{dom} L$; on the other hand, I do not assume that $\operatorname{dom} L$ has nonempty interior. □

Proof. The claim is trivially true if L takes the value $+\infty$ everywhere but at one point. Otherwise, $\operatorname{dom} L$ has a (nonempty, convex) *relative* interior.[13] Suppose that

$$\liminf_{\substack{x\to\bar{x}\\ x\in\operatorname{dom} L}} L(x) = \lim_{\substack{x\to\bar{x}\\ x\in\operatorname{dom} L}} L(x) = +\infty. \tag{17}$$

(This can only happen if $\bar{x}$ belongs to the relative boundary of $\overline{\operatorname{dom} L}$.) Let $\bar{\xi}$ be any point in the relative interior. (The line segment $[\bar{x}, \bar{\xi}]$ lies in $\overline{\operatorname{dom} L}$.) By [15, Corollary 7.5.1], and by uniqueness of the limit (17),

$$L(\bar{x}) = \lim_{\lambda\to 1} L(\lambda\bar{x} + (1-\lambda)\bar{\xi}) = +\infty = \liminf_{\substack{x\to\bar{x}\\ x\in\operatorname{dom} L}} L(x).$$

Suppose, on the other hand, that

$$\liminf_{\substack{x\to\bar{x}\\ x\in\operatorname{dom} L}} L(x) = M < +\infty.$$

Let $(x_n)_{n=1}^{\infty}$ be a sequence in $\operatorname{dom} L$ converging to $\bar{x}$ and such that $\lim_{n\to\infty} L(x_n) = M$. By convexity, for each n,

$$L(\bar{x}/2 + x_n/2) \le L(\bar{x})/2 + L(x_n)/2,$$

[13] The relative interior of a set $S \subseteq \mathbb{R}^d$ is the interior of S with respect to the topology of the smallest affine subspace of $\mathbb{R}^d$ that contains S; the relative boundary of S is its boundary with respect to the same topology. For instance, the smallest affine subspace of $\mathbb{R}^3$ containing a line segment $[\bar{x}_1, \bar{x}_2] \subset \mathbb{R}^3$ is the affine subspace ℓ, of dimension 1, generated by $\bar{x}_1$ and $\bar{x}_2$. The relative interior of $[\bar{x}_1, \bar{x}_2]$ is the segment $(\bar{x}_1, \bar{x}_2)$ (without the endpoints), and its relative boundary is the set $\{\bar{x}_1, \bar{x}_2\}$. See, e.g., [15] for further details.

and hence

$$M \leq \liminf_{n\to\infty} L(\bar{x}/2 + x_n/2) \leq \lim_{n\to\infty} L(\bar{x})/2 + L(x_n)/2 = L(\bar{x})/2 + M/2,$$
$$M \leq L(\bar{x}).$$

Since it also holds that $L(\bar{x}) \leq M$ by lower semicontinuity, $L(\bar{x}) = M$; this concludes the proof. ◻

LEMMA 8. *Suppose that a function $L : \mathbb{R}^d \to \overline{\mathbb{R}}$ is proper, convex, and lower semicontinuous. Then there exists a finite or countable set $S \subseteq \operatorname{dom} L$ such that for all $\bar{x} \in \overline{\operatorname{dom}}\, L$*

$$\liminf_{\substack{x\to\bar{x}\\ x\in S}} L(x) = L(\bar{x}).$$

Remark. From the standpoint of this paper, Lemmas 7 and 8 are only propaedeutic to the proof of Proposition 9; therefore their statements involve a function $L : \mathbb{R}^d \to \overline{\mathbb{R}}$. Nevertheless, to my understanding, they can be generalized without substantial changes at least to functions $L : V \to \overline{\mathbb{R}}$, where V is any separable Banach space. ◻

Proof. For any $i \in \mathbb{N}$, there exists a finite or countable covering $\mathcal{B}_i = \{B(c_j, 1/i)\}_{j=1}^{\infty}$ of $\overline{\operatorname{dom}}\, L$ made of open balls with centers c_j and radius $1/i$; let $\mathcal{B} = \bigcup_{i=1}^{\infty} \mathcal{B}_i$. From each open ball $B(c_j, 1/i) \in \mathcal{B}_i$ select a point $x_{ij} \in \operatorname{dom} L$ such that

$$\inf_{x\in B(c_j,1/i)} L(x) \leq L(x_{ij}) \leq \inf_{x\in B(c_j,1/i)} L(x) + 1/i$$

and form the set $S_i = \{x_{ij}\}_{j=1}^{\infty}$.[14] Let $S = \bigcup_{i=1}^{\infty} S_i$. The set $S \subseteq \operatorname{dom} L$ is everywhere dense in $\overline{\operatorname{dom}}\, L$, and at most countable.

Fix now $\bar{x} \in \overline{\operatorname{dom}}\, L$. By definition,

$$\liminf_{\substack{x\to\bar{x}\\ x\in \operatorname{dom} L}} L(x) = \lim_{n\to\infty} \inf_{x\in B(\bar{x},1/n)\cap \operatorname{dom} L\setminus\{\bar{x}\}} L(x).$$

Note that the clauses "$\cap \operatorname{dom} L$" and "$\setminus\{\bar{x}\}$" are both inessential, the first one because $L \equiv +\infty$ outside $\operatorname{dom} L$, and the second one because by Lemma 7 the limit inferior is $\leq L(\bar{x})$. So let me simplify notation:

$$\liminf_{\substack{x\to\bar{x}\\ x\in \operatorname{dom} L}} L(x) = \lim_{n\to\infty} \inf_{x\in B(\bar{x},1/n)} L(x). \tag{18}$$

For each $n \in \mathbb{N}$, by construction, there exists an open ball $B(c_n, r_n) \in \mathcal{B}$ such that $\bar{x} \in B(c_n, r_n)$ and $B(c_n, r_n) \subseteq B(\bar{x}, 1/n)$. The corresponding point $x_n \in S$ taken in the construction at the beginning of the proof satisfies

$$L(x_n) \geq \inf_{x\in B(c_n,r_n)} L(x) \geq \inf_{x\in B(\bar{x},1/n)} L(x); \tag{19}$$

on the other hand for each n there exists $h_n \in \mathbb{N}$ big enough such that $B(\bar{x}, 1/(n + h_n)) \subseteq B(c_n, r_n)$; by construction it follows that

$$L(x_n) \leq \inf_{x\in B(c_n,r_n)} L(x) + r_n \leq \inf_{x\in B(\bar{x},1/(n+h_n))} L(x) + 1/n, \tag{20}$$

[14] I know that an engineer invoking the axiom of choice may sound a bit booming, but... there it is.

because $r_n \le 1/n$. As $n \to \infty$, $x_n \to \bar{x}$, and both the right-hand sides of (19) and (20) converge to the limit inferior (18) (possibly to $+\infty$, with a slight abuse of terminology). Thus we have

$$\liminf_{\substack{x\to\bar{x}\\ x\in S}} L(x) \le \lim_{n\to\infty} L(x_n) = \liminf_{\substack{x\to\bar{x}\\ x\in \operatorname{dom} L}} L(x) \le \liminf_{\substack{x\to\bar{x}\\ x\in S}} L(x),$$

where the first inequality holds because $\{x_n\}_{n=1}^{\infty} \subseteq S$ and the second one holds because $S \subseteq \operatorname{dom} L$; hence

$$\liminf_{\substack{x\to\bar{x}\\ x\in S}} L(x) = \liminf_{\substack{x\to\bar{x}\\ x\in \operatorname{dom} L}} L(x),$$

and the claim follows by an application of Lemma 7. ☐

Remarks. (A) The only case in which the set S built in Lemma 8 is finite is when $\operatorname{dom} L$ contains only one point $\bar{x}$ (of course in this case $S = \{\bar{x}\}$): if $\operatorname{dom} L$ contains at least two points, then S must be countably infinite. (B) The statement "for all $x \in \mathcal{X}$, for almost all $\delta \in \Delta$, $f_\delta(x) \le \mathbf{L}(x)$" is trivial. Much less trivial is a statement like "for almost all $\delta \in \Delta$, for all $x \in \mathcal{X}$, $f_\delta(x) \le \mathbf{L}(x)$," because the clause "*almost* all" is preserved only by countable intersections, but $\mathcal{X} \subseteq \mathbb{R}^d$ is either trivial or uncountable, and hence the quantifiers "all x" and "almost all δ" cannot be interchanged freely. Of the second kind is the following proposition, that indeed relies essentially on the countability of S established by Lemma 8. ☐

PROPOSITION 9. *For all N, P^N-almost surely $\hat{f}_N(x) \le \mathbf{L}(x)$ for all $x \in \mathcal{X}$.*

Proof. The claim is trivial if $\mathbf{L} \equiv +\infty$, and hence suppose that $\mathbf{L}$ is proper. (It is also convex and lower semicontinuous by Propositions 5 and 6.) Since $\mathcal{X} \subseteq \mathbb{R}^d$, without loss of generality we can extend $\mathbf{L}$ to the whole of $\mathbb{R}^d$, letting $\mathbf{L}(x) = +\infty$ for all $x \in \mathbb{R}^d \setminus \mathcal{X}$. Consider now the countable, everywhere dense subset $S = \{x_n\}_{n\in\mathbb{N}} \subseteq \operatorname{dom} \mathbf{L}$ whose existence with respect to $\mathbf{L}$ is established by Lemma 8. Let

$$D_n = \{\delta \in \Delta \;:\; f_\delta(x_n) \le \mathbf{L}(x_n)\}, \qquad \bar{D} = \bigcap_{n=1}^{\infty} D_n;$$

since $\mathsf{P}[D_n] = 1$ for all $n \in \mathbb{N}$, also $\mathsf{P}\left[\bar{D}\right] = 1$. Fix N; for any choice of $\delta^{(1)}, \dots, \delta^{(N)} \in \bar{D}$, it holds that

$$\hat{f}_N(x_n) = \max_{i=1,\dots,N} f_{\delta^{(i)}}(x_n) \le \mathbf{L}(x_n) \quad \text{for all } x_n \in S.$$

Now let $\bar{x} \in \mathcal{X}$. If $\bar{x} \in \overline{\operatorname{dom}}\,\mathbf{L}$, the lower semicontinuity of $\hat{f}_N$ and Lemma 8 imply

$$\hat{f}_N(\bar{x}) \le \liminf_{x\to\bar{x}} \hat{f}_N(x) \le \liminf_{\substack{x\to\bar{x}\\ x\in S}} \hat{f}_N(x) \le \liminf_{\substack{x\to\bar{x}\\ x\in S}} \mathbf{L}(x) = \mathbf{L}(\bar{x}).$$

If instead $\bar{x} \notin \overline{\operatorname{dom}}\,\mathbf{L}$, then $\hat{f}_N(\bar{x}) \le \mathbf{L}(\bar{x}) = +\infty$ trivially. Since the choice of $\bar{x}$ is arbitrary, $\hat{f}_N(\bar{x}) \le \mathbf{L}(\bar{x})$ for all $\bar{x} \in \mathcal{X}$. This happens for all $(\delta^{(1)}, \dots, \delta^{(N)}) \in \bar{D}^N = \bar{D} \times \cdots \times \bar{D}$ (N times). But now, since $\delta^{(1)}, \dots, \delta^{(N)}$ are supposed to be *independent* random elements (or, which is the same, P^N is the product measure $\mathsf{P} \times \cdots \times \mathsf{P}$),

$$\mathsf{P}^N\left[\bar{D}^N\right] = \prod_{i=1}^{N} \mathsf{P}\left[\bar{D}\right] = 1,$$

and hence the uniform upper bound is almost sure; this proves the claim. ☐

4. Convergence of the solution (x^*_N, y^*_N). This section establishes another fundamental property of $\mathbf{L}$, namely, that it is coercive if Assumption 2 holds, two properties of the violation probability V, one probabilistic ($\mathsf{V}(x^*_N, y^*_N) \to 0$ almost surely) and one analytical (V is, in a sense, lower semicontinuous), and finally the main results of the paper: if Assumptions 1 and 2 hold, then $y^*_N \to \min \mathbf{L}$ almost surely, and if $x^{**} = \arg\min \mathbf{L}$ is unique, then $x^*_N \to x^{**}$ "almost surely" (Theorem 14); the case for $\mathbf{L} \equiv +\infty$ is covered by Theorem 15. The first thing I need to do is to assign a rigorous meaning to the clause "almost surely", that I have tacitly left vague from the beginning for the sake of readability.

To this purpose, turn now to consider the probability space of infinite sequences extracted independently from $(\Delta, \mathcal{F}, \mathsf{P})$, which I will denote $(\Delta^\infty, \mathcal{F}^\infty, \mathsf{P}^\infty)$. An element $\delta^\infty = (\delta^{(i)})_{i=1}^\infty \in \Delta^\infty$ is an infinite sequence of elements in Δ; $\mathcal{F}^\infty$ is the smallest σ-algebra of subsets of Δ^∞ containing the sets

$$\bar{\mathcal{E}}^{(N)} = \left\{ (\delta^{(1)}, \dots, \delta^{(N)}, \dots) \in \Delta^\infty \ : \ (\delta^{(1)}, \dots, \delta^{(N)}) \in \mathcal{E}^{(N)} \right\}$$

for all $N \in \mathbb{N}$ and $\mathcal{E}^{(N)} \in \mathcal{F}^N$; and P^∞ is a probability function such that

$$\mathsf{P}^\infty \left[\bar{\mathcal{E}}^{(N)} \right] = \mathsf{P}^N \left[\mathcal{E}^{(N)} \right]$$

for all $N \in \mathbb{N}$ and $\mathcal{E}^{(N)} \in \mathcal{F}^N$; since P^N, $N \in \mathbb{N}$, are all product measures, such probability exists and is unique in view of Ionescu-Tulcea's theorem [18, Theorem 2, p. 249]. Moreover, let $\bar{\mathcal{F}}^N = \{\bar{\mathcal{E}}^{(N)} \in \mathcal{F}^\infty : \mathcal{E}^{(N)} \in \mathcal{F}^{(N)}\}$; then $\left(\bar{\mathcal{F}}^{(N)}\right)_{N=1}^\infty$ is a filtration in $\mathcal{F}^\infty$. Fix an arbitrary $\bar{x} \in \mathcal{X}$, and if for a certain N the solution (x^*_N, y^*_N) does not exist, let by convention $(x^*_N, y^*_N) = (\bar{x}, -\infty)$;[15] with this convention, the sequence of solutions $((x^*_1, y^*_1), \dots, (x^*_N, y^*_N), \dots)$ is a stochastic process adapted to $\left(\bar{\mathcal{F}}^{(N)}\right)_{N=1}^\infty$, and P^∞ is compatible with P^N also in the following sense: for all $N \in \mathbb{N}$ and for any Borel set $\mathcal{B} \subseteq (\mathcal{X} \times \mathbb{R})^N$,

$$\mathsf{P}^\infty \left[\ \left\{ (x^*_i, y^*_i)_{i=1}^\infty \in (\mathcal{X} \times \mathbb{R})^\infty \ : \ (x^*_i, y^*_i)_{i=1}^N \in \mathcal{B} \right\} \ \right] = \mathsf{P}^N \left[(x^*_i, y^*_i)_{i=1}^N \in \mathcal{B} \right].$$

Consider now the random variable

$$\hat{\hat{N}} = \min \left\{ N \in \mathbb{N} \ : \ (x^*_N, y^*_N) \text{ exists with } y^*_N > -\infty \right\}.$$

$\hat{\hat{N}}$ is a stopping time with respect to the filtration $\left(\bar{\mathcal{F}}^{(N)}\right)_{N=1}^\infty$. The following lemma ensures that, under Assumption 2, $\hat{\hat{N}}$ is P^∞-almost surely finite.

Lemma 10. *If Assumption 2 holds, then P^∞-almost surely there exists $\hat{N} \in \mathbb{N}$ such that $\hat{f}_{\hat{N}}$ is coercive. For all $N \geq \hat{N}$, $\hat{f}_N$ is coercive and attains a finite minimum.*

Proof. Let $p = \mathsf{P}^{\bar{N}} \left[\hat{f}_{\bar{N}} \text{ is coercive} \right] > 0$. For $k = 0, 1, 2, \dots$, the sets

$$\mathcal{B}_k = \left\{ \max_{i = k\bar{N}+1, \dots, (k+1)\bar{N}} f_{\delta^{(i)}} \text{ is not coercive} \right\} \subseteq \Delta^{(k+1)\bar{N}}$$

[15] Here, $-\infty$ is just a placeholder to mark the nonexistence of a solution and to ensure that in this case $(x^*_N, y^*_N) \notin \mathcal{X} \times \mathbb{R}$; it does *not* mean that $\hat{f}_N$ is unbounded from below. Visualize δ with uniform density in $[-1, 1]$, $\mathcal{X} = \mathbb{R}$, and $f_\delta(x) = e^{\delta x}$; if $\delta^{(1)}, \dots, \delta^{(N)}$ are all positive, then $\inf_{x \in \mathcal{X}} \hat{f}_N(x) = 0$ is not attained.

form a sequence of *independent* events, each with probability $1-p$. Therefore

$$\begin{aligned}
&\mathsf{P}^{\infty}\left[\text{for all } k \in \mathbb{N}, \max_{i=k\bar{N}+1,\dots,(k+1)\bar{N}} f_{\delta^{(i)}} \text{ is not coercive}\right] \\
&= \lim_{K\to\infty} \mathsf{P}^{(K+1)\bar{N}}\left[\text{for all } k \le K, \max_{i=k\bar{N}+1,\dots,(k+1)\bar{N}} f_{\delta^{(i)}} \text{ is not coercive}\right] \\
&= \lim_{K\to\infty} \prod_{k=0}^{K} \mathsf{P}^{(k+1)\bar{N}}\left[\max_{i=k\bar{N}+1,\dots,(k+1)\bar{N}} f_{\delta^{(i)}} \text{ is not coercive}\right] \\
&= \lim_{K\to\infty} (1-p)^{K+1} = 0,
\end{aligned}$$

and hence P^{∞}-almost surely there exists k finite such that $\max_{i=k\bar{N}+1,\dots,(k+1)\bar{N}} f_{\delta^{(i)}}$ is coercive, and a fortiori $\hat{f}_{(k+1)\bar{N}} = \max_{i=1,\dots,(k+1)\bar{N}} f_{\delta^{(i)}}$ is coercive. The first claim follows letting $\hat{N} = (k+1)\bar{N}$. Then, for all $N \ge \hat{N}$, $\hat{f}_N$ is proper, lower semicontinuous, and coercive (because $\hat{f}_N(x) \ge \hat{f}_{\hat{N}}(x)$ for all $x \in \mathcal{X}$). Hence

$$y_N^* = \min_{x\in\mathcal{X}} \hat{f}_N(x) < +\infty \quad \text{and} \quad x_N^* = \arg\min_{x\in\mathcal{X}} \hat{f}_N(x)$$

exist by the Tonelli–Weierstrass theorem. ∎

PROPOSITION 11. *If Assumption* 2 *holds, then* $\mathbf{L}$ *is coercive.*

Proof. If $\mathbf{L}$ takes the only value $+\infty$, the claim is trivial, since any sublevel set of $\mathbf{L}$ is the empty set. Assume, therefore, that $\mathbf{L}$ is proper, and suppose for the sake of contradiction that it is not coercive. Then there exists $t \in \mathbb{R}$ such that the set $C = \{x \in \mathcal{X} : \mathbf{L}(x) \le t\}$ is nonempty and not compact. The set C is anyway closed, because $\mathbf{L}$ is lower semicontinuous [12, Proposition 2.2.5]; therefore, by the Heine–Borel theorem [16, Theorem 2.41], C is unbounded.

Proposition 9 ensures that, for all $N \in \mathbb{N}$,

$$\mathsf{P}^{\infty}\left[\hat{f}_N(x) \le \mathbf{L}(x) \text{ for all } x \in \mathcal{X}\right] = \mathsf{P}^{N}\left[\hat{f}_N(x) \le \mathbf{L}(x) \text{ for all } x \in \mathcal{X}\right] = 1.$$

Hence the event $\mathcal{D} \in \mathcal{F}^{\infty}$ defined as follows,

$$\mathcal{D} = \bigcap_{N=1}^{\infty} \left\{\hat{f}_N(x) \le \mathbf{L}(x) \text{ for all } x \in \mathcal{X}\right\},$$

has also probability 1. Therefore

$$\begin{aligned}
&\mathsf{P}^{\infty}\text{-almost surely, for all } N,\ \hat{f}_N(x) \le \mathbf{L}(x) \text{ for all } x \in \mathcal{X} \\
&\quad \Rightarrow \text{almost surely, for all } N,\ \{x \in \mathcal{X} : \hat{f}_N(x) \le t\} \supseteq \{x \in \mathcal{X} : \mathbf{L}(x) \le t\} = C \\
&\quad \Rightarrow \text{almost surely, for all } N,\ \{x \in \mathcal{X} : \hat{f}_N(x) \le t\} \text{ is unbounded.}
\end{aligned}$$

On the other hand, by Lemma 10, P^{∞}-almost surely there exists $\hat{N}$ finite such that $\hat{f}_{\hat{N}}$ is coercive, and hence P^{∞}-almost surely there exists $\hat{N}$ such that $\{x \in \mathcal{X} : \hat{f}_{\hat{N}}(x) \le t\}$ is compact. That $\{x \in \mathcal{X} : \hat{f}_{\hat{N}}(x) \le t\}$ is both compact and unbounded is a contradiction, stemming from the assumption that $\mathbf{L}$ is not coercive. ∎

The following two propositions establish an asymptotic property of $\mathsf{V}(x_N^*, y_N^*)$ and an analytic property of the function V (in essence, very similar to lower semicontinuity).

PROPOSITION 12.

$$\lim_{N\to\infty} \mathsf{V}(x_N^*, y_N^*) = 0 \quad \mathsf{P}^\infty\text{-almost surely.}$$

Proof. Fix $\varepsilon \in (0,1)$. By Lemma 10, P^∞-almost surely there exists $\hat{N} \in \mathbb{N}$ such that for all $N \geq \hat{N}$ the solution (x_N^*, y_N^*) exists. For all such N, by Theorem 2 (see (10)),

$$\begin{aligned}
\mathsf{P}^\infty\left[\mathsf{V}(x_N^*, y_N^*) > \varepsilon\right] &= \mathsf{P}^N\left[\mathsf{V}(x_N^*, y_N^*) > \varepsilon\right] \\
&\leq \sum_{k=0}^{d} \binom{N}{k} \varepsilon^k (1-\varepsilon)^{N-k} \\
&\leq \sum_{k=0}^{d} N^{d+1} (1-\varepsilon)^{N-d} \\
&= (d+1) N^{d+1} (1-\varepsilon)^{N-d}.
\end{aligned} \tag{21}$$

Since as $N \to \infty$

$$\frac{(d+1)(N+1)^{d+1}(1-\varepsilon)^{N+1-d}}{(d+1)N^{d+1}(1-\varepsilon)^{N-d}} = \left(1 + \frac{1}{N}\right)^{d+1} (1-\varepsilon) \to (1-\varepsilon) < 1,$$

by the ratio test

$$\sum_{N=\hat{N}}^{\infty} \mathsf{P}^\infty\left[\mathsf{V}(x_N^*, y_N^*) > \varepsilon\right] \leq \sum_{N=\hat{N}}^{\infty} (d+1) N^{d+1} (1-\varepsilon)^{N-d} < \infty.$$

Therefore, by the Borel–Cantelli lemma [18, p. 255],

$$\mathsf{P}^\infty\left[\mathsf{V}(x_N^*, y_N^*) > \varepsilon \text{ infinitely often}\right] = 0.$$

Actually this holds for all $\varepsilon \in (0,1]$, the case $\varepsilon = 1$ being trivial, and hence

$$\begin{aligned}
&\mathsf{P}^\infty\left[\text{there exists } h \in \mathbb{N} \text{ s.t. } \mathsf{V}(x_N^*, y_N^*) > 1/h \text{ infinitely often}\right] \\
&\quad = \mathsf{P}^\infty\left[\bigcup_{h=1}^{\infty} \left\{\mathsf{V}(x_N^*, y_N^*) > 1/h \text{ infinitely often}\right\}\right] \\
&\quad \leq \sum_{h=1}^{\infty} \mathsf{P}^\infty\left[\mathsf{V}(x_N^*, y_N^*) > 1/h \text{ infinitely often}\right] = 0.
\end{aligned}$$

Since the event {there exists $h \in \mathbb{N}$ such that $\mathsf{V}(x_N^*, y_N^*) > 1/h$ infinitely often} has probability 0, the complementary event {for all $h \in \mathbb{N}$ there exists $\bar{N} \in \mathbb{N}$ such that $0 \leq \mathsf{V}(x_N^*, y_N^*) \leq 1/h$ for all $N \geq \bar{N}$} has probability 1; the latter is included in the event {for any $\varepsilon' \in (0,1)$ there exist $h \in \mathbb{N}$ and $\bar{N} < \infty$ such that $0 \leq \mathsf{V}(x_N^*, y_N^*) \leq 1/h \leq \varepsilon'$ for all $N \geq \bar{N}$}, that is,

$$\left\{\delta^\infty \in \Delta^\infty \ : \ \lim_{N\to\infty} \mathsf{V}(x_N^*, y_N^*) = 0\right\}, \tag{22}$$

and hence (22) has also probability 1; this proves the claim. ∎

Remark. Equation (21) is the only point of this paper that makes use of the bound (10) provided by Theorem 2; the reader will of course notice that the second inequality is rather loose. Indeed I have chosen to present Theorem 2 in the introduction because that result provides the state-of-the-art bound for the scenario approach with convex constraints, but an inequality similar to (21), and equally sufficient for my purposes, would follow also from the pioneering result by Calafiore and Campi [3], which in the present context would read

$$\mathsf{P}^N\left[\mathsf{V}(x_N^*, y_N^*) > \varepsilon\right] \le \binom{N}{d+1}(1-\varepsilon)^{N-d-1}. \tag{23}$$ ◻

PROPOSITION 13. *Suppose that $(x_N)_{N=1}^{\infty}$ is a sequence in $\mathcal{X}$ converging to x_∞ as $N \to \infty$ and that $(y_N)_{N=1}^{\infty}$ is a nondecreasing sequence in $\mathbb{R}$ such that $y_N \le \bar{y} < +\infty$ for all $N \in \mathbb{N}$. Then*

$$\liminf_{N\to\infty} \mathsf{V}(x_N, y_N) \ge \mathsf{V}(x_\infty, \bar{y}).$$

In particular

$$\liminf_{N\to\infty} \mathsf{V}(x_N, y_N) \ge \mathsf{V}(x_\infty, y_\infty),$$

where $y_\infty = \lim_{N\to\infty} y_N$.

Proof. Consider random variables of the form $\mathbb{1}(f_\delta(x_N) > \bar{y})$. First note that, for all $\delta \in \Delta$ and $N \in \mathbb{N}$, $\mathbb{1}(f_\delta(x_N) > \bar{y}) \le \mathbb{1}(f_\delta(x_N) > y_N)$, since $y_N \le \bar{y}$.

Now fix $\delta \in \Delta$ and note that the indicator function can only take the values 0 and 1; therefore if $\liminf_{N\to\infty} \mathbb{1}(f_\delta(x_N) > \bar{y}) = 0$, it must hold that $f_\delta(x_N) \le \bar{y}$ for infinitely many N; then, by lower semicontinuity,

$$f_\delta(x_\infty) \le \liminf_{N\to\infty} f_\delta(x_N) \le \bar{y}.$$

By contraposition, if $f_\delta(x_\infty) > \bar{y}$, then $\liminf_{N\to\infty} \mathbb{1}(f_\delta(x_N) > \bar{y}) = 1$. It follows that

$$\mathbb{1}(f_\delta(x_\infty) > \bar{y}) \le \liminf_{N\to\infty} \mathbb{1}(f_\delta(x_N) > \bar{y}) \le \liminf_{N\to\infty} \mathbb{1}(f_\delta(x_N) > y_N).$$

Finally

$$\begin{aligned}
\liminf_{N\to\infty} \mathsf{V}(x_N, y_N) &= \liminf_{N\to\infty} \mathsf{P}\left[f_\delta(x_N) > y_N\right] \\
&= \liminf_{N\to\infty} \int_\Delta \mathbb{1}(f_\delta(x_N) > y_N)\, d\mathsf{P} \\
&\ge \int_\Delta \liminf_{N\to\infty} \mathbb{1}(f_\delta(x_N) > y_N)\, d\mathsf{P} \\
&\ge \int_\Delta \mathbb{1}(f_\delta(x_\infty) > \bar{y})\, d\mathsf{P} = \mathsf{V}(x_\infty, \bar{y}),
\end{aligned}$$

where the first inequality is due to Fatou's lemma [18, p. 187]. The second claim follows trivially letting $\bar{y} = y_\infty$. ◻

We are now ready to prove the main results of the paper.

THEOREM 14. *Suppose that Assumptions* 1 *and* 2 *hold and that* $\mathbf{L}$ *is proper. Then*

1. $\mathbf{L}$ *attains its minimum $y^{**} = \min_{x\in\mathcal{X}} \mathbf{L}(x) < +\infty$ at at least one point $x^{**} \in \mathcal{X}$;*

2. *irrespective of the uniqueness of x^{**}, it holds that*

$$\lim_{N\to\infty} y_N^* = y^{**} \quad \mathsf{P}^\infty\text{-almost surely;}$$

3. *if, moreover, x^{**} is unique, then*

$$\lim_{N\to\infty} x_N^* = x^{**} \quad \mathsf{P}^\infty\text{-almost surely.}$$

Proof. Since $\mathbf{L}$ is proper, lower semicontinuous, and also coercive by Proposition 11, by the Tonelli–Weierstrass theorem there exists in $\mathcal{X}$ at least one minimum point x^{**} of $\mathbf{L}$ attaining a finite minimum $y^{**} = \mathbf{L}(x^{**}) < +\infty$. This proves point 1.

To prove point 2 first note that, by Lemma 10, there exists a set $\mathcal{D}_1 \subseteq \Delta^\infty$ with probability 1 such that, for all $\delta^\infty \in \mathcal{D}_1$, there exists $\hat{N} \in \mathbb{N}$ such that $\hat{f}_{\hat{N}}$ is coercive. Moreover, by Proposition 12, there exists a set $\mathcal{D}_2 \subseteq \Delta^\infty$ with probability 1 such that, for all $\delta^\infty \in \mathcal{D}_2$, $\mathsf{V}(x_N^*, y_N^*) \to 0$ as $N \to \infty$. And by Proposition 9 the set

$$\begin{aligned}\mathcal{D}_3 &= \Big\{\text{for all } N \in \mathbb{N},\ \hat{f}_N(x) \le \mathbf{L}(x) \text{ for all } x \in \mathcal{X}\Big\}\\ &= \bigcap_{N=1}^{\infty} \Big\{\hat{f}_N(x) \le \mathbf{L}(x) \text{ for all } x \in \mathcal{X}\Big\}\end{aligned}$$

has probability 1. Let $\mathcal{D} = \mathcal{D}_1 \cap \mathcal{D}_2 \cap \mathcal{D}_3$ (hence also $\mathsf{P}^\infty[\mathcal{D}] = 1$).

Fix a $\delta^\infty \in \mathcal{D}$ and the corresponding $\hat{N}$. For $N \ge \hat{N}$, the $\hat{f}_N$'s form a nondecreasing sequence of coercive functions, all bounded from above by $\mathbf{L}$, and their minima y_N^* form a nondecreasing sequence of real numbers bounded from above by y^{**}, which therefore has a limit y_∞^*. Let $C = \{x \in \mathcal{X} : \hat{f}_{\hat{N}}(x) \le y^{**}\}$. C is nonempty because it contains at least the minimum point x^{**}, closed because $\hat{f}_{\hat{N}}$ is lower semicontinuous, and bounded because $\hat{f}_{\hat{N}}$ is coercive. Thus, C is nonempty and compact. For all $N \ge \hat{N}$, since $\hat{f}_N(x) \ge \hat{f}_{\hat{N}}(x)$ for all $x \in \mathcal{X}$, it holds that $\{x \in \mathcal{X} : \hat{f}_N(x) \le y^{**}\} \subseteq C$. Therefore $x_N^* \in C$ for all $N \ge \hat{N}$, and since C is compact there exists a converging subsequence $(x_{N_k}^*)_{k=1}^\infty$ of the sequence $(x_N^*)_{N=\hat{N}}^\infty$. Let $x_\infty^* = \lim_{k\to\infty} x_{N_k}^*$; of course it holds also that $\lim_{k\to\infty} y_{N_k}^* = y_\infty^*$. Now

$$\begin{aligned}\mathsf{P}[f_\delta(x_\infty^*) > y_\infty^*] &= \mathsf{V}(x_\infty^*, y_\infty^*)\\ &\le \lim_{k\to\infty} \mathsf{V}(x_{N_k}^*, y_{N_k}^*) \qquad \text{(Proposition 13)}\\ &= 0 \qquad (\text{because } \delta^\infty \in \mathcal{D}_2).\end{aligned}$$

Hence $\mathsf{P}[f_\delta(x_\infty^*) > y_\infty^*] = 0$, and consequently

$$\mathbf{L}(x_\infty^*) = \inf\{y \in \mathbb{R} \ :\ \mathsf{P}[f_\delta(x_\infty^*) > y] = 0\} \le y_\infty^*. \tag{24}$$

Since $y^{**} = \min \mathbf{L}$, it follows that $y_\infty^* \ge y^{**}$; on the other hand $y_\infty^* \le y^{**}$ by construction. Therefore $\lim_{N\to\infty} y_N^* = y^{**}$; and since the choice of $\delta^\infty \in \mathcal{D}$ was arbitrary and $\mathsf{P}^\infty[\mathcal{D}] = 1$, this proves point 2.

To prove point 3, suppose that x^{**} is unique. Fix $\delta^\infty \in \mathcal{D}$ and the corresponding $\hat{N}$, consider the nonempty compact set C as in the proof of point 2, and recall that $x_N^* \in C$ for all $N \ge \hat{N}$. Suppose also, for the sake of contradiction, that x_N^* does not converge to x^{**}. Then there exists $\eta > 0$ such that the set $\{x \in \mathcal{X} : \|x - x^{**}\| \ge \eta\}$ contains infinitely many terms of the sequence $(x_N^*)_{N=\hat{N}}^\infty$. Since $\{\|x - x^{**}\| \ge \eta\}$

is closed, the intersection $K = \{\|x - x^{**}\| \geq \eta\} \cap C$ is nonempty and compact, and hence from the sequence $(x_N^*)_{N=\hat{N}}^{\infty}$ we can extract a converging subsequence $(x_{N_k}^*)_{k=1}^{\infty}$ of elements in K and let $x_\infty^* = \lim_{k\to\infty} x_{N_k}^*$. (Again, it also holds that $\lim_{k\to\infty} y_{N_k}^* = y_\infty^*$.) But then $x_\infty^* \in K$, and hence $\|x_\infty^* - x^{**}\| \geq \eta$ so that $x_\infty^* \neq x^{**}$. Finally,

$$\begin{aligned} y^{**} = \mathbf{L}(x^{**}) &< \mathbf{L}(x_\infty^*) \qquad \text{(uniqueness of the minimum point)} \\ &\leq y_\infty^* \qquad \text{(inequality (24)).} \end{aligned}$$

That $\lim_{N\to\infty} y_N^* > y^{**}$ is a contradiction (indeed $y_N^* \leq y^{**}$ for all N), stemming from the assumption that x_N^* does not converge to x^{**}. Hence $\lim_{N\to\infty} x_N^* = x^{**}$ for all $\delta^\infty \in \mathcal{D}$; this concludes the proof of the theorem. ∎

THEOREM 15. *Suppose that Assumptions* 1 *and* 2 *hold and that* $\mathbf{L}$ *takes the only value* $+\infty$. *Then*

$$\lim_{N\to\infty} y_N^* = +\infty \quad \mathsf{P}^\infty\text{-almost surely.}$$

Proof. Construct the set $\mathcal{D} \subseteq \Delta^\infty$, with probability 1, exactly as in the proof of Theorem 14, and fix $\delta^\infty \in \mathcal{D}$ and the corresponding $\hat{N}$. For $N \geq \hat{N}$, the $\hat{f}_N$'s form a nondecreasing sequence of coercive functions, and their minima y_N^* form a nondecreasing sequence of real numbers. For the sake of contradiction suppose that, for all N, $y_N^* \leq \bar{y}$ for a certain $\bar{y} \in \mathbb{R}$. Let $C = \{x \in \mathcal{X} : \hat{f}_{\hat{N}}(x) \leq \bar{y}\}$; the set C is nonempty and compact, and for all $N \geq \hat{N}$, since $\hat{f}_N(x) \geq \hat{f}_{\hat{N}}(x)$ for all $x \in \mathcal{X}$, it holds that $\{x \in \mathcal{X} : \hat{f}_N(x) \leq \bar{y}\} \subseteq C$. It follows that $x_N^* \in C$ for all $N \geq \hat{N}$, and due to the compactness of C there exists a converging subsequence $(x_{N_k}^*)_{k=1}^{\infty}$ of the sequence $(x_N^*)_{N=\hat{N}}^{\infty}$. Let $x_\infty^* = \lim_{k\to\infty} x_{N_k}^*$; we have

$$\begin{aligned} \mathsf{P}\left[f_\delta(x_\infty^*) > \bar{y}\right] &= \mathsf{V}(x_\infty^*, \bar{y}) \\ &\leq \lim_{k\to\infty} \mathsf{V}(x_{N_k}^*, y_{N_k}^*) \qquad \text{(Proposition 13)} \\ &= 0 \qquad \text{(because } \delta^\infty \in \mathcal{D}). \end{aligned}$$

Hence $\mathsf{P}\left[f_\delta(x_\infty^*) > \bar{y}\right] = 0$, and consequently

$$\mathbf{L}(x_\infty^*) = \inf\{y \in \mathbb{R} \;:\; \mathsf{P}\left[f_\delta(x_\infty^*) > y\right] = 0\} \leq \bar{y}.$$

This is a contradiction ($\mathbf{L}(x_\infty^*) = +\infty$), stemming from the assumption that the sequence $(y_N^*)_{N=\hat{N}}^{\infty}$ is bounded from above; hence $\lim_{N\to\infty} y_N^* = +\infty$. Since $\delta^\infty \in \mathcal{D}$ is arbitrary and $\mathsf{P}^\infty[\mathcal{D}] = 1$, this proves the theorem. ∎

5. Back to the scenario approach. This section is meant to apply Theorems 14 and 15 to the scenario problem (5) and to the essential robust problem (7) introduced in Section 1; let therefore $\mathcal{X} = \Theta \subseteq \mathbb{R}^d$ and $x = \theta$. Recall from section 2.1 that the solution of (5) is equivalent to the minimization of $\hat{f}_N(\theta) = \max_{i=1\cdots N} \; f_{\delta^{(i)}}(\theta)$, where

$$f_\delta(\theta) = c^\top\theta + \mathcal{I}_{\Theta_\delta}(\theta) = \begin{cases} c^\top\theta & \text{if } \theta \in \Theta_\delta, \\ +\infty & \text{otherwise,} \end{cases} \tag{25}$$

and denote

$$\begin{aligned} \theta_N^* &= \arg\min_{\theta\in\Theta} \hat{f}_N(\theta), \\ y_N^* &= \min_{\theta\in\Theta} \hat{f}_N(\theta) = c^\top\theta_N^*. \end{aligned}$$

Assumption 1 requires that for all $\delta \in \Delta$ the function f_δ is convex and lower semicontinuous. Since the term $c^\top \theta$ is linear, this is the same as requiring that $\mathcal{I}_{\Theta_\delta}$ is convex and lower semicontinuous; and since Θ_δ is the 0-sublevel set of $\mathcal{I}_{\Theta_\delta}$ this is, in turn, equivalent to the requirement that Θ_δ is convex and closed. Hence in the present setting Assumption 1 can be restated as follows.

ASSUMPTION 3. *The set $\Theta \subseteq \mathbb{R}^d$ is convex and closed; for all $\delta \in \Delta$, Θ_δ is convex and closed. For all $\theta \in \Theta$, the event $\{\theta \in \Theta_\delta\}$ belongs to $\mathcal{F}$.*

Assumption 2 requires that $\hat{f}_N$ is almost surely proper. Since $\hat{f}_N(\theta) < +\infty$ if and only if θ belongs to all the sets $\Theta_{\delta^{(i)}}$, this is equivalent to the requirement that the intersection of these sets is nonempty. Moreover, the assumption requires that the probability of the event $\{\hat{f}_{\bar{N}}$ is coercive$\}$ is nonzero for a certain $\bar{N}$. $\hat{f}_{\bar{N}}$ is coercive when its t-sublevel set is either empty (when $t < \min \hat{f}_{\bar{N}}$) or *bounded* (equivalently, *compact*, since $\hat{f}_{\bar{N}}$ is lower semicontinuous and all its sublevel sets are closed). Thus, for the problem at hand, Assumption 2 can be restated as follows.

ASSUMPTION 4. *For all $N \in \mathbb{N}$, P^N-almost surely $\bigcap_{i=1}^N \Theta_{\delta^{(i)}} \neq \varnothing$. Moreover, there exists $\bar{N} \in \mathbb{N}$ such that the event*

$$\left\{ \text{for some } t \in \mathbb{R}, \text{ the set } \bigcap_{i=1}^{\bar{N}} \Theta_{\delta^{(i)}} \cap \{c^\top \theta \le t\} \text{ is nonempty and bounded} \right\}$$

has nonzero probability.

Next, I need to translate in the language of problems (5)–(7) the distinction between when $\mathbf{L}$ is proper and when it takes the only value $+\infty$. This is done in the following lemma.

LEMMA 16. *If the functions f_δ, $\delta \in \Delta$, are defined as in* (25), *then $\mathbf{L}$ is proper if and only if there exists $\bar{\theta} \in \Theta$ such that $\mathsf{P}\left[\bar{\theta} \in \Theta_\delta\right] = 1$, i.e., if and only if the essential robust problem* (7),

$$\begin{aligned} &\min_{\theta \in \Theta} \; c^\top \theta \\ &\text{subject to} \quad \mathsf{P}\left[\theta \in \Theta_\delta\right] = 1, \end{aligned}$$

is feasible. If $\mathbf{L}$ is proper, the solution of problem (7) *is* $\arg\min_{\theta \in \Theta} \mathbf{L}(\theta)$.

Proof. For any $y \in \mathbb{R}$, $f_\delta(\theta) > y$ if and only if $c^\top \theta > y$ or $\theta \notin \Theta_\delta$. Therefore $\mathsf{P}\left[f_\delta(\theta) > y\right] = \mathsf{P}\left[c^\top \theta > y \text{ or } \theta \notin \Theta_\delta\right]$, and we recover three inequalities:

$$\begin{aligned} &\mathsf{P}\left[f_\delta(\theta) > y\right] \le \mathbb{1}(c^\top \theta > y) + \mathsf{P}\left[\theta \notin \Theta_\delta\right], \\ &\mathsf{P}\left[f_\delta(\theta) > y\right] \ge \mathbb{1}(c^\top \theta > y), \\ &\mathsf{P}\left[f_\delta(\theta) > y\right] \ge \mathsf{P}\left[\theta \notin \Theta_\delta\right]. \end{aligned} \tag{26}$$

If $\mathsf{P}\left[\theta \notin \Theta_\delta\right] = 0$ (equivalently $\mathsf{P}\left[\theta \in \Theta_\delta\right] = 1$), then the first two inequalities in (26) imply $\mathsf{P}\left[f_\delta(\theta) > y\right] = \mathbb{1}(c^\top \theta > y)$, and hence $\mathsf{P}\left[f_\delta(\theta) > y\right] = 0$ if and only if $c^\top \theta \le y$ and, by definition,

$$\mathbf{L}(\theta) = \inf\left\{y \in \mathbb{R} \;:\; \mathsf{P}\left[f_\delta(\theta) > y\right] = 0\right\} = \inf\left\{y \in \mathbb{R} \;:\; c^\top \theta \le y\right\} = c^\top \theta.$$

If instead $\mathsf{P}\left[\theta \notin \Theta_\delta\right] > 0$, then the third inequality in (26) implies $\mathsf{P}\left[f_\delta(\theta) > y\right] > 0$ for all $y \in \mathbb{R}$ and, again by definition,

$$\mathbf{L}(\theta) = \inf\left\{y \in \mathbb{R} \;:\; \mathsf{P}\left[f_\delta(\theta) > y\right] = 0\right\} = \inf \varnothing = +\infty.$$

Summing up,

$$\mathbf{L}(\theta) = \begin{cases} c^\top \theta & \text{if } \mathsf{P}\left[\theta \in \Theta_\delta\right] = 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Thus, if there exists $\bar{\theta} \in \Theta$ such that $\mathsf{P}\left[\bar{\theta} \in \Theta_\delta\right] = 1$, then $\mathbf{L}(\bar{\theta}) = c^\top \bar{\theta}$ and $\mathbf{L}$ is proper; otherwise $\mathbf{L}$ takes the only value $+\infty$, and this proves the first claim. From the last observation it is also clear that

$$\begin{aligned} &\min_{\theta \in \Theta} \; c^\top \theta \quad \text{subject to} \quad \mathsf{P}\left[\theta \in \Theta_\delta\right] = 1 \\ = &\min_{\theta \in \Theta} \; \mathbf{L}(\theta) \quad \text{subject to} \quad \mathbf{L}(\theta) < +\infty \\ = &\min_{\theta \in \Theta} \; \mathbf{L}(\theta) \quad (\text{if } \mathbf{L} \text{ is proper}); \end{aligned}$$

the second claim follows immediately. □

Remark. At first sight, it might seem that the problem of finding $\arg\min_{\theta \in \Theta} \mathbf{L}(\theta)$, that is. the unconstrained minimization of a convex and lower semicontinuous function, being conceptually simpler than the constrained minimization in problem (7), could be simpler also computationally, i.e., more convenient to solve by means of numerical methods. This is in general *false*. Letting aside the choice of a method to minimize $\mathbf{L}$, the true difficulty here is the *computation* of $\mathbf{L}$: indeed deciding whether $\mathbf{L}(\theta) = c^\top \theta$ or $\mathbf{L}(\theta) = +\infty$ for a given θ may be an intractable problem.[16] The reason of the computational difficulty is made explicit by the following example, proposed by Ben-Tal and Nemirovski in [1, section 3.2.2, p. 787]. Consider the robust problem

$$\begin{aligned} &\min_{\theta \in \mathbb{R}^n} \; c^\top \theta \\ &\text{subject to} \quad \theta^\top \mathrm{diag}(\delta)^2 \theta \le 1 \text{ for all } \delta \in \Delta, \end{aligned} \tag{27}$$

where Δ is a parallelotope in $\mathbb{R}^n$ centered at the origin; assume that Δ is equipped with, say, a uniform distribution. The corresponding essential robust problem is

$$\begin{aligned} &\min_{\theta \in \mathbb{R}^n} \; c^\top \theta \\ &\text{subject to} \quad \mathsf{P}\left[\theta^\top \mathrm{diag}(\delta)^2 \theta \le 1\right] = 1. \end{aligned} \tag{28}$$

Both (27) and (28) are feasible (the former because $\theta = 0$ satisfies the constraint $\theta^\top \mathrm{diag}(\delta)^2 \theta \le 1$ for all $\delta \in \Delta$, and the latter because $\mathsf{P}[\Delta] = 1$); actually, the two problems are the same.[17] But, for instance, deciding the robust feasibility of the point $\bar{\theta} = [1\ 1\ 1\ \cdots\ 1]^\top$, that is, deciding whether or not $\mathbf{L}(\bar{\theta})$ is finite, amounts to verifying that $\|\delta\|_2 \le 1$ for all $\delta \in \Delta$, and this is known to be a NP-hard problem. (Refer to [1] for more details.) □

From the standpoint of section 1 the following theorem may also be thought of as the main result of the paper. Nevertheless, in view of Lemma 16 and of the previous discussion, it is clearly just a restatement of Theorems 14 and 15, and therefore I leave it without proof.

[16] The same story can be told about the computation of the convex and lower semicontinuous function $\mathbf{S}$, whose minimization corresponds to the solution of the robust problem (1).

[17] Problems (27) and (28) coincide because the mapping $\delta \mapsto \theta^\top \mathrm{diag}(\delta)^2 \theta$ is continuous for all θ and, assuming uniform distribution, the support of P is the whole Δ: this implies $\mathbf{L} = \mathbf{S}$, as has already been mentioned in a footnote.

THEOREM 17. *Suppose that Assumptions* 3 *and* 4 *hold. If the essential robust problem* (7) *is feasible, then*

1. *problem* (7) *admits a solution* θ^{**}*;*
2. *irrespective of the uniqueness of* θ^{**}*, it holds that*

$$\lim_{N\to\infty} c^\top \theta_N^* = c^\top \theta^{**} \quad \mathsf{P}^\infty\text{-almost surely;}$$

3. *if, moreover,* θ^{**} *is unique, then*

$$\lim_{N\to\infty} \theta_N^* = \theta^{**} \quad \mathsf{P}^\infty\text{-almost surely.}$$

If instead problem (7) *is infeasible, then*

$$\lim_{N\to\infty} c^\top \theta_N^* = +\infty \quad \mathsf{P}^\infty\text{-almost surely.} \qquad \square$$

To conclude the discussion, I will try to provide a bit more insight about the possible infeasibility of the essential robust problem. It has already been observed, in a comment below Definition 3, that the requirement that $\hat{f}_N$ is almost surely proper (Assumption 2) does not exclude that $\mathbf{L} \equiv +\infty$ so that $\lim_{N\to\infty} \hat{f}_N(\theta) = +\infty$ almost surely for all $\theta \in \Theta$. In pretty much the same way, $\bigcap_{i=1}^N \Theta_{\delta^{(i)}} \neq \varnothing$ does not exclude that $\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} = \varnothing$; indeed this is what happens almost surely in the "bad" case where problem (7) is infeasible and $c^\top \theta_N^* \to +\infty$. This connection is clarified by the following result.

PROPOSITION 18. *Under the hypotheses of Theorem* 17*, problem* (7) *is infeasible if and only if*

$$\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} = \varnothing \quad \mathsf{P}^\infty\text{-almost surely.}$$

A necessary condition for this to hold is that $\bigcap_{i=1}^N \Theta_{\delta^{(i)}}$ *is* P^N*-almost surely unbounded for all* N*.*

Proof. Suppose that problem (7) is infeasible. Consider a $\delta^\infty \in \Delta^\infty$ such that $\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}}$ is not empty. For any such δ^∞ there exists $\bar\theta \in \bigcap_{i=1}^\infty \Theta_{\delta^{(i)}}$, and it holds that $\lim_{N\to\infty} y_N^* \le c^\top \bar\theta < +\infty$. Hence, by the last claim of Theorem 17, the set of all such δ^∞ has probability 0. Vice versa, suppose that almost surely $\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} = \varnothing$ and, for the sake of contradiction, that there exists $\bar\theta \in \Theta$ such that $\mathsf{P}\left[\bar\theta \in \Theta_\delta\right] = 1$. Then, by independence,

$$\mathsf{P}^\infty\left[\bar\theta \in \bigcap_{i=1}^\infty \Theta_{\delta^{(i)}}\right] = \lim_{N\to\infty} \mathsf{P}^N\left[\bar\theta \in \bigcap_{i=1}^N \Theta_{\delta^{(i)}}\right] = \lim_{N\to\infty} \prod_{i=1}^N \mathsf{P}\left[\bar\theta \in \Theta_{\delta^{(i)}}\right] = 1,$$

i.e., almost surely $\bar\theta \in \bigcap_{i=1}^\infty \Theta_{\delta^{(i)}}$, which is a contradiction. Hence such a $\bar\theta$ does not exist, and problem (7) is infeasible; this proves the first claim. To prove the second claim suppose again that almost surely $\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} = \varnothing$ and, for the sake of contradiction, that $\bigcap_{i=1}^N \Theta_{\delta^{(i)}}$ is bounded with nonzero probability for a certain N. Let $C_k = \bigcap_{i=1}^{(k+1)N} \Theta_{\delta^{(i)}}$. The sets $(C_k)_{k=0}^\infty$ form a nonincreasing sequence $\cdots C_{k-1} \supseteq C_k \supseteq C_{k+1} \cdots$ of nonempty closed sets; but since

$$\mathsf{P}^N\left[\bigcap_{i=1}^N \Theta_{\delta^{(kN+i)}} \text{ is bounded}\right] = \mathsf{P}^N\left[\bigcap_{i=1}^N \Theta_{\delta^{(i)}} \text{ is bounded}\right] > 0 \quad \text{for all } k \in \mathbb{N},$$

P^∞-almost surely there exists $\hat{k} \in \mathbb{N}$ such that $\bigcap_{i=1}^N \Theta_{\delta^{(\hat{k}N+i)}}$ is bounded. Hence, for $k \geq \hat{k}$ the intersections C_k are nonempty and *compact*. It follows [16, Corollary to Theorem 2.36] that $\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} = \bigcap_{k=0}^\infty C_k \neq \varnothing$, P^∞-*almost surely*. This is a contradiction stemming from the assumption that $\mathsf{P}^N[\bigcap_{i=1}^N \Theta_{\delta^{(i)}}$ is bounded$] > 0$; hence $\mathsf{P}^N[\bigcap_{i=1}^N \Theta_{\delta^{(i)}}$ is bounded$] = 0$ for all N, and this concludes the proof. ∎

Remarks. (A) If problem (7) is infeasible, then, of course, problem (1) is also infeasible, because $\mathsf{P}[\Delta] = 1$. (B) The main point of the second claim of Proposition 18 is to help establish feasibility by contraposition: for example, if $\mathsf{P}[\Theta_\delta$ is bounded$] > 0$ or, more generally, if there exist finitely many events, say, $A_1, \ldots, A_{\bar{N}} \subseteq \Delta$, each with positive probability, such that for any choice of $\delta^{(i)} \in A_i$, $i = 1, \ldots, \bar{N}$, the set $\Theta_{\delta^{(1)}} \cap \cdots \cap \Theta_{\delta^{(\bar{N})}}$ is bounded, then problem (7) must be feasible. Moreover, if these events exist the second part of Assumption 4 is automatically enforced, because if $\bigcap_{i=1}^{\bar{N}} \Theta_{\delta^{(i)}}$ is bounded there always exists $t \in \mathbb{R}$ such that $\bigcap_{i=1}^{\bar{N}} \Theta_{\delta^{(i)}} \subseteq \{c^\top \theta \leq t\}$. The message of Proposition 18 and of this remark is illustrated in the following, final, example. ∎

Example 6. Let $p \in (0,1)$, $\Theta = \mathbb{R}^2$, $\theta = (x_1, x_2)$, $c^\top\theta = x_1 + x_2$, and consider three cases. (A) Suppose that $\delta = (r, \gamma)$, where r is a binary variable taking values $1, 2$ with equal probability $1/2$ and γ is a random variable independent of r with geometric distribution $\mathsf{P}[\gamma = n] = p(1-p)^n$. Let moreover

$$\Theta_\delta = \begin{cases} \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 \geq \gamma\right\} & \text{if } r = 1, \\ \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_2 \geq \gamma\right\} & \text{if } r = 2. \end{cases}$$

Letting $G_{N,1} = \max_{i=1,\ldots,N:\ r^{(i)}=1} \gamma^{(i)}$ and $G_{N,2} = \max_{i=1,\ldots,N:\ r^{(i)}=2} \gamma^{(i)}$, it holds that $\bigcap_{i=1}^N \Theta_{\delta^{(i)}} = \{x_1 \geq G_{N,1},\ x_2 \geq G_{N,2}\}$. In this case, P^∞-almost surely, as $N \to \infty$ both $G_{N,1}$ and $G_{N,2}$ tend to $+\infty$, the minimum $c^\top\theta_N^* = G_{N,1} + G_{N,2}$ also tends to $+\infty$, and, as claimed by Proposition 18, $\bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} = \varnothing$; and of course $\bigcap_{i=1}^N \Theta_{\delta^{(i)}}$ is P^N-almost surely unbounded for all N. A pictorial view of this case is shown in Figure 6(a).

(B) Suppose that $\delta = (r, \gamma)$ as before, but this time let

$$\Theta_\delta = \begin{cases} \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 \geq -\frac{1}{\gamma+1}\right\} & \text{if } r = 1, \\ \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_2 \geq -\frac{1}{\gamma+1}\right\} & \text{if } r = 2. \end{cases}$$

(See Figure 6(b).) As before $\bigcap_{i=1}^N \Theta_{\delta^{(i)}}$ is P^N-almost surely unbounded for all N, but now $c^\top\theta_N^* = -1/(G_{N,1}+1) - 1/(G_{N,2}+1)$ tends almost surely to 0 (i.e., the minimum attained by problem (7)), and $\varnothing \neq \bigcap_{i=1}^\infty \Theta_{\delta^{(i)}} =$ the first quadrant of the x_1, x_2 plane. This point shows that the unboundedness of $\bigcap_{i=1}^N \Theta_{\delta^{(i)}}$ for all N is only necessary for the infeasibility of (7), not sufficient.

(C) Finally, suppose that r is a *ternary* variable taking values $1, 2, 3$ with equal probability $1/3$, $\delta = (r, \gamma)$ as before, and

$$\Theta_\delta = \begin{cases} \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 \geq -\frac{1}{\gamma+1}\right\} & \text{if } r = 1, \\ \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_2 \geq -\frac{1}{\gamma+1}\right\} & \text{if } r = 2, \\ \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ x_1 \leq 5,\ x_2 \leq 5\right\} & \text{if } r = 3. \end{cases}$$

(See Figure 6(c).) It is immediate to check that $\bigcap_{i=1}^N \Theta_{\delta^{(i)}} \neq \varnothing$ for all N. Moreover, although Θ_δ is unbounded for every $\delta \in \Delta$, the events $A_1 = \{\delta \in \Delta : r = 1\}$,

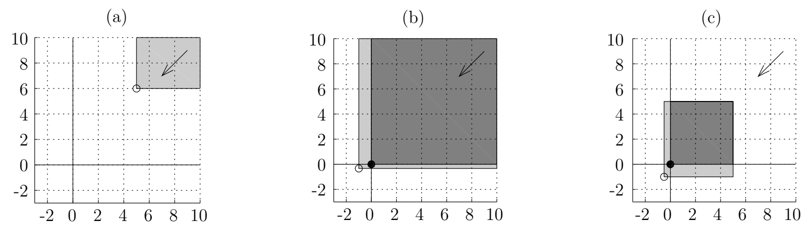

FIG. 6. *Instances of finite-sample problems for each case in Example* 6. *Light gray:* $\bigcap_{i=1}^{N} \Theta_{\delta^{(i)}}$*; dark gray:* $\bigcap_{i=1}^{\infty} \Theta_{\delta^{(i)}}$*;* $\circ = \theta_N^*$*;* $\bullet$ = *essential robust solution.* (a) *The essential robust problem is infeasible.* (b) *The essential robust problem is feasible but* $\bigcap_{i=1}^{N} \Theta_{\delta^{(i)}}$ *is* P^N*-almost surely unbounded for all* N. (c) *The essential robust problem is feasible and* $\bigcap_{i=1}^{N} \Theta_{\delta^{(i)}}$ *is compact.*

$A_2 = \{\delta \in \Delta : r = 2\}$, and $A_3 = \{\delta \in \Delta : r = 3\}$, each with positive probability, are such that if $\delta^{(i)} \in A_i$, $i = 1, 2, 3$, then the set $\Theta_{\delta^{(1)}} \cap \Theta_{\delta^{(2)}} \cap \Theta_{\delta^{(3)}}$ is always bounded. According to the second claim of Proposition 18, this is enough to establish the feasibility of problem (7). P^∞-almost surely, the solution θ_N^* converges to the origin (i.e., the solution of (7)) and $\bigcap_{i=1}^{\infty} \Theta_{\delta^{(i)}} = [0, 5] \times [0, 5]$. ◻

6. Conclusions. In this paper I have shown that the solution of a convex scenario program, subject to independent and identically distributed constraints that, almost surely, sooner or later confine it to a compact set (coercivity), converges almost surely to the solution of a suitably defined *essential* robust problem. Future work may be dedicated

- to searching for weak enough conditions such that the scenario solution converges to the solution of the *robust* problem, i.e., nice sufficient conditions to ensure that $\mathbf{L}(x) = \mathbf{S}(x)$ at least at the minimum points $x = x^{**}$ of $\mathbf{L}$;
- to weakening the convexity hypothesis. On one hand, it should not be difficult to show that the convergence proved here holds also for certain families of nonconvex problems, for example, the one considered in [13, section IV]. On the other hand, the scenario approach can be applied also to nonconvex problems, albeit with a different a posteriori assessment method (see, e.g., [6], [8]), but to my knowledge no general a priori bounds like (10) or (23), powerful enough to exploit the Borel–Cantelli lemma (proof of Proposition 12), exist without the assumptions of convexity and finite-dimensionality ($\mathcal{X} \subseteq \mathbb{R}^d$). Convexity has also been used to prove that, for all N, almost surely $\hat{f}_N \leq \mathbf{L}$ (Proposition 9), but I conjecture that a *uniform* bound is not necessary to establish the main results.

Acknowledgments. I am indebted to Rinaldo Colombo, to Augusto Ferrante, and to my colleagues Giorgio Arici, Marco Campi, Algo Carè, Marco Dalai, and Renata Mansini for their fruitful suggestions about the paper, and particularly to Simone Garatti for helping me lift a technical assumption about Δ. I extend my gratitude to the anonymous reviewers for their constructive and encouraging comments.

REFERENCES

[1] A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res., 23 (1998), pp. 769–805.
[2] D. P. Bertsekas, *Convex Optimization Theory*, Athena Scientific, Belmont, MA, 2009.
[3] G. C. Calafiore and M. C. Campi, *The scenario approach to robust control design*, IEEE Trans. Automat. Control, (2006), pp. 742–753.

[4] M. C. Campi and S. Garatti, *The exact feasibility of randomized solutions of uncertain convex programs*, SIAM J. Optim., 19 (2008), pp. 1211–1230.
[5] M. C. Campi and S. Garatti, *A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality*, J. Optim. Theory Appl., 148 (2011), pp. 257–280.
[6] M. C. Campi and S. Garatti, *Wait-and-judge scenario optimization*, Math. Program., Sprnger, New York, 2016, https://doi.org/10.1007/s10107-016-1056-9.
[7] M. C. Campi, S. Garatti, and M. Prandini, *The scenario approach for systems and control design*, Annu. Rev. Control, 33 (2009), pp. 149–157.
[8] M. C. Campi, S. Garatti, and F. A. Ramponi, *Non-convex scenario optimization with application to system identification*, in Proceedings of the 54th IEEE Conference on Decision and Control, Osaka, Japan, 2015.
[9] A. Carè, S. Garatti, and M. C. Campi, *Scenario min-max optimization and the risk of empirical costs*, SIAM J. Optim., 25 (2015), pp. 2061–2080.
[10] T. Kanamori and A. Takeda, *Worst-case violation of sampled convex programs for optimization with uncertainty*, J. Optim. Theory Appl., 152 (2012), pp. 171–197.
[11] A. J. Kleywegt, A. Shapiro, and T. Homem-de-Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim., 12 (2001), pp. 479–502.
[12] R. Lucchetti, *Convexity and Well-Posed Problems*, Springer, New York, 2006.
[13] P. Mohajerin Esfahani, T. Sutter, and J. Lygeros, *Performance bounds for the scenario approach and an extension to a class of non-convex programs*, IEEE Trans. Automat. Control, 60 (2015), pp. 46–58.
[14] A. Prékopa, *Contributions to the theory of stochastic programming*, Math. Program., 4 (1973), pp. 202–221.
[15] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1997.
[16] W. Rudin, *Principles of Mathematical Analysis*, 3rd intl. ed., McGraw-Hill, New York, 1976.
[17] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming—Modeling and Theory*, SIAM, Philadelphia, 2009.
[18] A. N. Shiryaev, *Probability*, Springer, New York, 1996.