Asymptotic Consistency for Nonconvex Risk-Averse Stochastic Optimization with Infinite Dimensional Decision Spaces

Johannes Milz

Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany, milz@ma.tum.de

Thomas M. Surowiec

FB12 Mathematik und Informatik, Philipps-Universität Marburg, Marburg, Germany, surowiec@mathematik.uni-marburg.de

Optimal values and solutions of empirical approximations of stochastic optimization problems can be viewed as statistical estimators of their true values. From this perspective, it is important to understand the asymptotic behavior of these estimators as the sample size goes to infinity, which is both of theoretical as well as practical interest. This area of study has a long tradition in stochastic programming. However, the literature is lacking consistency analysis for problems in which the decision variables are taken from an infinite-dimensional space, which arise in optimal control, scientific machine learning, and statistical estimation. By exploiting the typical problem structures found in these applications that give rise to hidden norm compactness properties for solution sets, we prove consistency results for nonconvex risk-averse stochastic optimization problems formulated in infinite-dimensional space. The proof is based on several crucial results from the theory of variational convergence. The theoretical results are demonstrated for several important problem classes arising in the literature.

Key words: asymptotic consistency, empirical approximation, sample average approximation, Monte Carlo sampling, risk-averse optimization, PDE-constrained optimization, uncertainty quantification, stochastic programming

MSC2000 subject classification: 90C15, 90C06, 62F12, 35Q93, 49M41, 49J52

OR/MS subject classification: Primary: ; secondary:

History: Received July 22, 2022

1. Introduction   The asymptotic behavior of empirical approximations is a central point of study in optimization under uncertainty. There is a long tradition going back to the fundamental contributions [29, 20, 57, 52, 58, 31, 59, 47, 60, 50, 48, 38, 64, 51]. These works have since given rise to standard derivation techniques for problems with finite-dimensional decision spaces. There are in essence three main techniques used to obtain asymptotic statements. The first possibility uses epi-convergence of sample-based approximations of objective functions over compact sets and therefore draws from powerful statements in the theory of variational convergence. The second type of method employs a uniform law of large numbers for sample-based approximations of objective functions. Finally, asymptotic statements can also be derived from stability estimates for optimal values and solutions with respect to probability semimetrics. This requires, amongst other things, that the class of integrands in the objective constitutes a P-uniformity class for the semimetric in question.

Given a general stochastic optimization problem

$$\min_{z \in Z_{\mathrm{ad}}} \mathbb{E}_{\mathbb{P}}[F(z)], \tag{1}$$

an empirical approximation would take the form

$$\min_{z \in Z_{\mathrm{ad}}} \mathbb{E}_{\mathbb{P}_N}[F(z)], \tag{2}$$

where the original probability measure $\mathbb{P}$ is replaced by a (sequence of) typically discrete approximation(s) $\mathbb{P}_N$ for $N \in \mathbb{N}$. For example, the probability measure $\mathbb{P}_N$ could be an empirical probability measure associated with a random sample of size N from $\mathbb{P}$. This is a common approach often referred to as "sample average approximation" (SAA), see e.g., [32, 63]. A data-driven viewpoint can be drawn from machine learning in which (1) represents the "population risk minimization" problem and (2) the corresponding "empirical risk minimization" problem. Here, the underlying probability measure of the data $\mathbb{P}$ is typically unknown. It is therefore of interest to understand the behavior of solutions in the big data limit (as $N \to \infty$).

The main questions can be easily stated: Do the optimal values and solution sets of (2) converge to their "true" counterparts for (1) as N passes to infinity and what is the strongest form of stochastic convergence that can be guaranteed? If we treat the N-dependent objects as statistical estimators of the true values and seek to prove at least convergence in probability, then these are questions of consistency, cf. [61].

Motivated by recent advances in partial differential equation (PDE)-constrained optimization under uncertainty [22, 35], scientific machine learning [9, 46], nonconvex stochastic programming [40, 49, 16], and statistical estimation [54, 55, 41], we provide such consistency results for stochastic optimization problems in which the decision variables z may be taken in an infinite-dimensional space Z. We will consider more general "risk-averse" problems in which the expectation $\mathbb{E}_{\mathbb{P}}$ is allowed to be replaced by certain classes of convex risk functionals $\mathcal{R}$. And as it is often lacking in the application areas mentioned above, we do not assume convexity of the integrand F. For consistency results on finite-dimensional risk-averse stochastic optimization problems, we refer the reader to [61, 63].

From an abstract perspective, we consider stochastic optimization problems of the type

$$\min_{z \in Z_{\mathrm{ad}}} \mathcal{R}[F(z)] + \wp(z). \tag{3}$$

Here, Z_{ad} is typically a closed convex subset of an infinite-dimensional space, e.g., $L^2(D)$; $\wp$ is a deterministic convex cost function; F is a random integrand that typically depends on the solution of a differential equation subject to random inputs; and $\mathcal{R}$ is a convex functional that acts as a numerical surrogate for our risk preference, e.g., a convex combination of $\mathbb{E}_{\mathbb{P}}[X]$ and a semideviation $\mathbb{E}_{\mathbb{P}}[\max\{0, X - \mathbb{E}[X]\}]$.

Despite the past successes in consistency analysis listed above, there is a major difficulty in extending the finite-dimensional arguments to the infinite-dimensional setting. In order to use both the epigraphical as well as the uniform law of large numbers approaches, we need an appropriately defined *norm* compact set that contains both the approximate N-dependent solutions as well as true solutions. It is not enough for the feasible set to be closed and bounded. For example, the simple set of pointwise bilateral constraints

$$Z_{\mathrm{ad}} := \left\{ z \in L^2(0,1) \colon 0 \leq z(x) \leq 1 \text{ for a.e. } x \in (0,1) \right\}$$

is weakly sequentially compact in $L^2(0,1)$, but not norm compact. The literature is not void of results for infinite-dimensional problems. However, the stability statements developed in [28, 53]

and the large deviation-type bounds derived in [44, 43] have only been demonstrated for strongly convex risk-neutral problems. While it may be possible to extend some of these results to a risk-averse setting, it appears rather challenging to obtain statements about the consistency of minimizers without strong convexity. In the recent preprint [42], consistency results for optimal values and solutions are established for risk-neutral PDE-constrained optimization using a uniform law of large numbers.

The paper is structured as follows. In Section 2, we introduce the basic notation, assumptions, and several preliminary results necessary for the remaining parts of the text. Afterwards, in Section 3, we present our main result. Finally, the utility of the main consistency result is demonstrated for several problem classes in Section 4.

2. Notation, Assumptions, and Preliminary Results   We introduce several concepts, notation and assumptions that are required in the text.

2.1. Probability and Function Spaces   Throughout the text, all spaces are defined over the real numbers $\mathbb{R}$ and metric spaces are equipped with their Borel σ-field. Let Ξ be a complete separable metric space, $\mathcal{A}$ the associated Borel σ-algebra, and $\mathbb{P}: \mathcal{A} \to [0,1]$ a probability measure. The triple $(\Xi, \mathcal{A}, \mathbb{P})$ is always assumed to be a complete probability space. Throughout the manuscript, $(\Omega, \mathcal{F}, P)$ is a probability space.

If Υ is a Banach space, then its topological dual space is denoted by Υ^*. If Υ is reflexive, we identify its bi-dual $(\Upsilon^*)^*$ with Υ. Throughout the text, we will use $p \in [1, \infty)$ for a general integrability exponent. In the application section, we will consider problems involving random partial differential equations. These require several function spaces. The underlying physical domain $D \subset \mathbb{R}^d$ with $d \in \{1, 2, 3\}$ will always be an open bounded Lipschitz domain.

For a Banach space $(V, \|\cdot\|_V)$ we will denote the Lebesgue–Bochner space $L^p(\Xi, \mathcal{A}, \mathbb{P}; V)$ of all strongly $\mathcal{F}$-measurable V-valued functions by

$$L^p(\Xi, \mathcal{A}, \mathbb{P}; V) = \{u : \Xi \to V : u \text{ strongly } \mathcal{F}\text{-measurable and } \|u\|_{L^p(\Xi, \mathcal{A}, \mathbb{P}; V)} < \infty\}$$

endowed with the natural norms $\|u\|_{L^p(\Xi, \mathcal{A}, \mathbb{P}; V)} = (\mathbb{E}_{\mathbb{P}}[\|u\|_V^p])^{1/p}$ for $p \in [1, \infty)$ and for bounded fields: $\|u\|_{L^\infty(\Xi, \mathcal{A}, \mathbb{P}; V)} = \mathbb{P}\text{-ess}\sup_{\xi \in \Xi} \|u(\xi)\|_V$. In the event that $V = \mathbb{R}$, we simply write $L^p(\Omega, \mathcal{F}, P)$. For the PDE applications, we use $L^p(D)$ to denote the usual Lebesgue space of p-integrable (or essentially bounded) functions over D. For more details on Lebesgue–Bochner spaces, we refer the reader to [27, Chapter III]. We denote convergence in the norm by $\to$ and weak convergence by $\rightharpoonup$.

Given two random variables $X_1, X_2 \in L^p(\Omega, \mathcal{F}, P)$ for $p \in [1, \infty)$, we say that X_1 and X_2 are *distributionally equivalent* with respect to P if $P(X_1 \le t) = P(X_2 \le t)$ for all $t \in \mathbb{R}$. A functional $\rho: L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is said to be *law invariant* with respect to P if for all distributionally equivalent random variables $X_1, X_2 \in L^p(\Omega, \mathcal{F}, P)$ we have $\rho(X_1) = \rho(X_2)$. In this setting, it therefore makes sense to use the (abuse of) notation $\rho(F_X)$, where $F_X(t) = P(X \le t)$ with $t \in \mathbb{R}$ as opposed to $\rho(X)$. We caution that this does not mean we redefine the function ρ over a space of càdlàg functions.

2.2. Convex Analysis and Several Key Functionals   Given a Banach space V, the (effective) domain of an extended real-valued function $f: V \to (-\infty, \infty]$, will be denoted by $\mathrm{dom}(f) \coloneqq \{x \in V : f(x) < \infty\}$. We typically exclude convex functions that take the value $-\infty$. For $f: V \to (-\infty, \infty]$ and $\varepsilon > 0$, $x_\varepsilon \in V$ is a ε-minimizer of f provided $\inf_{v \in V} f(v)$ is *finite* and $f(x_\varepsilon) \le \inf_{v \in V} f(v) + \varepsilon$. The ε-solution set $(\varepsilon \ge 0)$ is then the set $\mathcal{S}^\varepsilon \coloneqq \{x \in V : f(x) \le \inf_{v \in V} f(v) + \varepsilon\}$. We use the convention $\mathcal{S} = \mathcal{S}^0$.

Let Υ be a normed space. For $x \in \Gamma \subset \Upsilon$ and $\Psi \subset \Upsilon$, we define

$$\mathrm{dist}(\Gamma, \Psi) = \inf_{y \in \Psi} \|x - y\|_\Upsilon \quad \text{and} \quad \mathbb{D}(\Gamma, \Psi) = \sup_{x \in \Gamma} \mathrm{dist}(x, \Psi). \tag{4}$$

We recall that a Banach space V has the *Radon–Riesz (Kadec–Klee) property* if $v_k \to v$ whenever $(v^k) \subset V$ is a sequence with $v^k \rightharpoonup v \in V$ and $\|v^k\|_V \to \|v\|_V$ as $k \to \infty$. More generally, we will say that a function $\varphi : V \to [0, \infty)$ is an *R-function* if it is convex and continuous, and if $v_k \to v$ as $k \to \infty$ whenever $(v^k) \subset V$ is a sequence with $v^k \rightharpoonup v \in V$ and $\varphi(v^k) \to \varphi(v)$ as $k \to \infty$. If V is a reflexive Banach space, then there exists an R-function on V [13, p. 154]. Notions related to that of an R-function are available in the literature, such as functions having the Kadec property and strongly rotund functions [13, 14] (see also [30]).

As the following fact demonstrates, the class of R-functions is rather large and includes, e.g., typical cost functions and regularizers used in PDE-constrained optimization.

LEMMA 1. *Let V be a Banach space. If $\wp : [0, \infty) \to [0, \infty)$ is convex and strictly increasing and $\varphi : V \to [0, \infty)$ is an R-function, then $\wp \circ \varphi$ is an R-function. In particular, if V has the Radon–Riesz property, then $\wp \circ \| \cdot \|_V$ is an R-function.*

Proof The function $\wp \circ \varphi$ is convex and continuous. Let $v_k \rightharpoonup v$ and $\wp(\varphi(v_k)) \to \wp(\varphi(v))$. Since $\wp$ is strictly increasing on $[0, \infty)$, it has a continuous inverse. Hence $\varphi(v_k) \to \varphi(v)$.

For a Banach space V and $(\Xi, \mathcal{A}, \mathbb{P})$ as above, $f : V \times \Xi \to (-\infty, \infty]$ is said to be *random lower semicontinuous* provided f is jointly measurable (with respect to the tensor-product σ-algebra of Borel σ-algebras on V and $\mathcal{A}$) and $f(\cdot, \xi)$ is lower semicontinuous for every $\xi \in \Xi$. If Υ_1 and Υ_2 are metric spaces, then $G : \Upsilon_1 \times \Xi \to \Upsilon_2$ is a *Carathéodory mapping* provided $G(v, \cdot)$ is measurable for all $v \in \Upsilon_1$ and $G(\cdot, \xi)$ is continuous for all $\xi \in \Xi$.

Finally, there are many concepts of risk measures in the literature. We will work with the following with further refinements as needed in the text below. Let $\rho : L^p(\Omega, \mathcal{F}, P) \to (-\infty, \infty]$. We consider the following conditions on the functional ρ.

(R1) Convexity. For all $X, Y \in L^p(\Omega, \mathcal{F}, P)$ and $\lambda \in (0, 1)$, we have $\rho(\lambda X + (1 - \lambda)Y) \leq \lambda \rho(X) + (1 - \lambda)\rho(Y)$.

(R2) Monotonicity. For all $X, Y \in L^p(\Omega, \mathcal{F}, P)$ such that $X \leq Y$ w.p. 1, we have $\rho(X) \leq \rho(Y)$.

(R3) Translation equivariance. If $X \in L^p(\Omega, \mathcal{F}, P)$ and C is a degenerate random variable with $C \equiv c$ w.p. 1 for some $c \in \mathbb{R}$, then $\rho(X + C) = \rho(X) + c$.

(R4) Positive homogeneity. If $X \in L^p(\Omega, \mathcal{F}, P)$ and $\gamma > 0$, then $\rho(\gamma X) = \gamma \rho(X)$.

The risk measure ρ is called *convex* if it satisfies (R1)–(R3) and it is referred to as *coherent* if it satisfies (R1)–(R4), see [2], [21], and in particular [63, p. 231].

2.3. Epiconvergence and Weak Inf-Compactness   Variational convergence, in particular (Mosco-)epiconvergence play a central role in consistency analysis. We provide here the necessary definitions and results from the literature. In addition, we prove several new results that are tailored to applications involving PDEs with random inputs.

We recall the notions of epiconvergence and Mosco-epiconvergence [4].

DEFINITION 1 (EPICONVERGENCE).   Let V be a complete metric space. Let $\phi_k : V \to (-\infty, \infty]$ be a sequence and let $\phi : V \to (-\infty, \infty]$ be a function. The sequence (ϕ_k) *epiconverges* to ϕ if for each $v \in V$

1. and each $(v_k) \subset V$ with $v_k \to v$ as $k \to \infty$, $\liminf_{k \to \infty} \phi_k(v_k) \geq \phi(v)$, and

2. there exists $(v_k) \subset V$ with $v_k \to v$ as $k \to \infty$ such that $\limsup_{k \to \infty} \phi_k(v_k) \leq \phi(v)$.

In many instances in infinite-dimensional optimization, especially the calculus of variations, optimal control, and PDE-constrained optimization we are forced to work with weaker topologies in the context of variational convergence. If the underlying space is a reflexive Banach space, then we may appeal to epiconvergence in the sense of Mosco, which was introduced in [45].

DEFINITION 2 (MOSCO-EPICONVERGENCE).   Let V be a reflexive Banach space and let $V_0 \subset V$ be a closed, convex set. Let $\phi_k : V_0 \to (-\infty, \infty]$ be a sequence and let $\phi : V_0 \to (-\infty, \infty]$ be a function. The sequence (ϕ_k) *Mosco-epiconverges* to ϕ if for each $v \in V_0$

1. and each $(v_k) \subset V_0$ with $v_k \rightharpoonup v$ as $k \to \infty$, $\liminf_{k\to\infty} \phi_k(v_k) \geq \phi(v)$, and
2. there exists $(v_k) \subset V_0$ with $v_k \to v$ such that $\limsup_{k\to\infty} \phi_k(v_k) \leq \phi(v)$.

In the definition of Mosco-epiconvergence, we allow for the sequence (ϕ_k) and the epi-limit ϕ to be defined on a convex, closed subset of a reflexive Banach space. This allows us to model constraints without the need for indicator functions. We will see below in Theorem 1 that this variation on the original definition leaves the crucial implications of Mosco-epiconvergence intact. In other words, Theorem 1 provides conditions sufficient for consistency of optimal values of Mosco-epiconvergent objective functions; compare with [3, Thm. 1.10], [17, Thm. 5.3], and [14, Thm. 6.2.8], for example.

THEOREM 1.   *Let V be a reflexive Banach space and let $V_0 \subset V$ be a closed, convex set. Suppose that $h_k \colon V_0 \to (-\infty, \infty]$ Mosco-epiconverges to $h \colon V_0 \to (-\infty, \infty]$. Let $(v^k) \subset V_0$ and $(\varepsilon_k) \subset [0, \infty)$ be sequences such that $\varepsilon_k \to 0^+$ and for each $k \in \mathbb{N}$, let v^k satisfy*

$$h_k(v^k) \leq \inf_{v \in V_0} h_k(v) + \varepsilon_k.$$

If $(v^k)_K$ is a subsequence of (v^k) such that $v^k \rightharpoonup \bar{v}$ as $K \ni k \to \infty$, then
1. $\bar{v} \in V_0$,
2. $h(\bar{v}) = \inf_{v \in V_0} h(v)$,
3. $\inf_{v \in V_0} h_k(v) \to \inf_{v \in V_0} h(v)$ *as $K \ni k \to \infty$,*
4. $h_k(v^k) \to h(\bar{v})$ *as $K \ni k \to \infty$.*

Proof Since (h_k) Mosco-epiconverges to h on V_0, it epiconverges to h, where V_0 may be understood as a complete metric space using the norm topology. Hence

$$\limsup_{k\to\infty} \inf_{v \in V_0} h_k(v) \leq \inf_{v \in V_0} h(v); \tag{5}$$

see, e.g., [3, Props. 1.14 and 2.9]. Since $(v^k) \subset V_0$ and V_0 is weakly sequentially closed, we have $\bar{v} \in V_0$. Then Mosco-epiconvergence ensures

$$\liminf_{K \ni k\to\infty} \inf_{v \in V_0} h_k(v) = \liminf_{K \ni k\to\infty} [\varepsilon_k + \inf_{v \in V_0} h_k(v)] \geq \liminf_{K \ni k\to\infty} h_k(v^k) \geq h(\bar{v}).$$

Combined with (5), we find that $h(\bar{v}) = \inf_{v \in V_0} h(v)$ and $\inf_{v \in V_0} h_k(v) \to \inf_{v \in V_0} h(v)$ as $K \ni k \to \infty$. The assertion $h_k(v^k) \to h(\bar{v})$ as $K \ni k \to \infty$ is implied by the above derivations and

$$\limsup_{K \ni k\to\infty} h_k(v^k) \leq \limsup_{K \ni k\to\infty} [\varepsilon_k + \inf_{v \in V_0} h_k(v)] = \limsup_{K \ni k\to\infty} \inf_{v \in V_0} h_k(v) \leq \limsup_{k\to\infty} \inf_{v \in V_0} h_k(v).$$

Proposition 1 demonstrates a weak compactness property of approximate minimizers to "regularized" optimization problems with Mosco-epiconvergent objective functions. Let Z be a reflexive Banach space, let $Z_{\mathrm{ad}} \subset Z$ be a closed, convex set, and let $f_k, f \colon Z_{\mathrm{ad}} \to (-\infty, \infty]$. Furthermore, let $\varphi \colon Z \to [0, \infty)$ be a convex, continuous function. We define the optimal values

$$\mathfrak{m}_k^* := \inf_{z \in Z_{\mathrm{ad}}} \{ f_k(z) + \varphi(z) \} \quad \text{and} \quad \mathfrak{m}^* := \inf_{z \in Z_{\mathrm{ad}}} \{ f(z) + \varphi(z) \} \tag{6}$$

and the solution sets

$$\mathcal{S}_k^{\varepsilon_k} := \{ z \in Z_{\mathrm{ad}} \colon f_k(z) + \varphi(z) \leq \mathfrak{m}_k^* + \varepsilon_k \} \quad \text{and} \quad \mathcal{S} := \{ z \in Z_{\mathrm{ad}} \colon f(z) + \varphi(z) = \mathfrak{m}^* \}.$$

The optimal values in (6) are somewhat related to evaluations of certain epi-sums [4].

PROPOSITION 1. *Let Z be a reflexive Banach space, let $Z_{\mathrm{ad}} \subset Z$ be a closed, convex set, let $\varphi : Z \to [0, \infty)$ be a convex, continuous function, and let $Z_0 \subset Z$ be bounded. Suppose that $f_k : Z_{\mathrm{ad}} \to (-\infty, \infty]$ Mosco-epiconverges to $f : Z_{\mathrm{ad}} \to (-\infty, \infty]$. Let $(\varepsilon_k) \subset [0, \infty)$ be a sequence with $\varepsilon_k \to 0^+$. Suppose that $\mathcal{S} \neq \emptyset$ and that for all $k \in \mathbb{N}$,*

$$\mathcal{S}_k^{\varepsilon_k} \subset Z_0 \quad and \quad \mathcal{S}_k^{\varepsilon_k} \neq \emptyset.$$

If (z^k) is a sequence with $z^k \in \mathcal{S}_k^{\varepsilon_k}$ for all $k \in \mathbb{N}$ and $(z^k)_K$ is a subsequence of (z^k), then $(z^k)_K$ has a further subsequence $(z^k)_{K_1}$ converging weakly to some $\bar{z} \in \mathcal{S}$ and $\varphi(z^k) \to \varphi(\bar{z})$ as $K_1 \ni k \to \infty$.

Proof Since $(z^k)_K \subset Z_{\mathrm{ad}}$, $(z^k)_K \subset Z_0$, Z_0 is bounded, and Z_{ad} is closed and convex, $(z^k)_K$ has a further subsequence $(z^k)_{K_1}$ such that $z^k \rightharpoonup \bar{z} \in Z_{\mathrm{ad}}$ as $K_1 \ni k \to \infty$ [12, Thms. 2.23 and 2.28]. Since $\bar{z} \in Z_{\mathrm{ad}}$, the Mosco-epiconvergence of (f_k) to f ensures the existence of a sequence $(\tilde{z}^k) \subset Z_{\mathrm{ad}}$ such that $\tilde{z}^k \to \bar{z} \in Z_{\mathrm{ad}}$ as $k \to \infty$ and $\limsup_{k \to \infty} f_k(\tilde{z}^k) \leq f(\bar{z})$. Since $\tilde{z}^k \to \bar{z}$ implies $\tilde{z}^k \rightharpoonup \bar{z}$, we have $\lim_{k \to \infty} f_k(\tilde{z}^k) = f(\bar{z})$. Since $\tilde{z}^k \in \mathcal{S}_k^{\varepsilon_k}$, we have for all $k \in \mathbb{N}$,

$$f_k(z^k) + \varphi(z^k) \leq f_k(\tilde{z}^k) + \varphi(\tilde{z}^k) + \varepsilon_k. \tag{7}$$

Since (f_k) Mosco-epiconverges to f, we have $f(\bar{z}) \leq \liminf_{K_1 \ni k \to \infty} f_k(z^k)$. Combined with the fact that φ is continuous and

$$\liminf_{K_1 \ni k \to \infty} f_k(z^k) + \limsup_{K_1 \ni k \to \infty} \varphi(z^k) \leq \limsup_{K_1 \ni k \to \infty} f_k(z^k) + \varphi(z^k),$$

the estimate (7) ensures

$$f(\bar{z}) + \limsup_{K_1 \ni k \to \infty} \varphi(z^k) \leq \limsup_{K_1 \ni k \to \infty} f_k(\tilde{z}^k) + \varphi(\tilde{z}^k) + \varepsilon_k \leq \limsup_{k \to \infty} f_k(\tilde{z}^k) + \varphi(\tilde{z}^k) + \varepsilon_k$$
$$= \lim_{k \to \infty} f_k(\tilde{z}^k) + \varphi(\tilde{z}^k) + \varepsilon_k = f(\bar{z}) + \varphi(\bar{z}).$$

Since $z^k \rightharpoonup \bar{z}$ as $K_1 \ni k \to \infty$, $\mathcal{S} \neq \emptyset$, and (f_k) Mosco-epiconverges to f, Theorem 1 ensures $\bar{z} \in \mathcal{S}$. Since $\bar{z} \in \mathcal{S}$, we have $f(\bar{z}) \in \mathbb{R}$. Thus $\limsup_{K_1 \ni k \to \infty} \varphi(z^k) \leq \varphi(\bar{z})$. Since φ is convex and continuous, it is weakly lower semicontinuous. Combined with $z^k \rightharpoonup \bar{z}$ as $K_1 \ni k \to \infty$, we have $\varphi(z^k) \to \varphi(\bar{z})$ as $K_1 \ni k \to \infty$.

While the sum of an Mosco-epiconvergent sequence and a convex, continuous function Mosco-epiconverge, Proposition 1 allows us to draw further conclusions about the minimizers to composite optimization problems defined by sums of Mosco-epiconvergent and convex, continuous functions than a direct application of the "sum rule." For example, if φ is an R-function, then the sequence $(z^k)_{K_1}$ considered in Proposition 1 converges strongly to an element of $\mathcal{S}$.

COROLLARY 1. *If the hypotheses of Proposition 1 hold true and φ is an R-function, then each subsequence of (z^k) has a further subsequence converging strongly to an element of $\mathcal{S}$.*

REMARK 1. If Z_{ad} is bounded, then we can choose $Z_0 = Z_{\mathrm{ad}}$ in Proposition 1. The condition $\mathcal{S}_k^{\varepsilon_k} \subset Z_0$ for all $k \in \mathbb{N}$ in Proposition 1 is related to a "weak inf-compactness" condition, provided that Z_0 is also convex and bounded. In this case, Z_0 is weakly (sequentially) compact. Instead of requiring $\mathcal{S}_k^{\varepsilon_k} \subset Z_0$ for all $k \in \mathbb{N}$, we could require for some $\gamma \in \mathbb{R}$ and for all $k \in \mathbb{N}$,

$$\emptyset \neq \{ z \in Z_{\mathrm{ad}} : f_k(z) + \varphi(z) \leq \gamma \} \subset Z_0. \tag{8}$$

The level set condition (8) ensures that $\mathcal{S}_k^{\varepsilon_k}$ is nonempty, provided that f_k is weakly lower semi-continuous. In case that Z_0 is norm compact, the condition (8) has been used, for example, in Theorem 2.1 in [38] to establish consistency properties for infinite-dimensional stochastic programs. If $\sup_{k \in \mathbb{N}} \mathfrak{m}_k^* < \infty$, $\gamma > \sup_{k \in \mathbb{N}} \mathfrak{m}_k^*$ and for all $k \in \mathbb{N}$,

$$\{ z \in Z : f_k(z) + \varphi(z) \leq \gamma \} \subset Z_0,$$

then $\mathcal{S}_k^{\varepsilon_k} \subset Z_0$ for all sufficiently large $k \in \mathbb{N}$ since we eventually have $\sup_{k \in \mathbb{N}} \mathfrak{m}_k^* + \varepsilon_k \leq \gamma$.

COROLLARY 2. *If the hypotheses of Proposition 1 hold, then $\mathfrak{m}_k^* \to \mathfrak{m}^*$ as $k \to \infty$. If furthermore φ is an R-function, then $\mathbb{D}(\mathcal{S}_k^{\varepsilon_k}, \mathcal{S}) \to 0$ as $k \to \infty$.*

Proof Let $z^k \in \mathcal{S}_k^{\varepsilon_k}$ for each $k \in \mathbb{N}$. The hypotheses ensure that (f_k) Mosco-epiconverges to f. Let $(\mathfrak{m}_k^*)_K$ be a subsequence of $(\mathfrak{m}_k^*)$. Proposition 1 ensures that $(z^k)_K$ has a further subsequence $(z^k)_{K_1}$ that weakly converges to some element in $\mathcal{S}$. Combined with Theorem 1, we find that $\mathfrak{m}_k^* \to \mathfrak{m}^*$ as $K_1 \ni k \to \infty$. Since $\mathcal{S}$ is nonempty, $\mathfrak{m}^* \in \mathbb{R}$. Putting together the pieces, we have shown that each subsequence of $(\mathfrak{m}_k^*)$ has a further subsequence converging to $\mathfrak{m}^*$. Hence $\mathfrak{m}_k^* \to \mathfrak{m}^*$ as $k \to \infty$.

It must still be shown that $\mathbb{D}(\mathcal{S}_k^{\varepsilon_k}, \mathcal{S}) \to 0$ as $k \to \infty$. Since $\mathcal{S}_k^{\varepsilon_k} \subset Z_0$ and $\mathcal{S} \subset Z$ are nonempty, and Z_0 is bounded, we have $\mathbb{D}(\mathcal{S}_k^{\varepsilon_k}, \mathcal{S}) \leq \mathbb{D}(Z_0, \mathcal{S}) < \infty$ for all $k \in \mathbb{N}$. Let us define the sequence $\varrho_k := \mathbb{D}(\mathcal{S}_k^{\varepsilon_k}, \mathcal{S})$. Let $(\varrho_k)_K$ be a subsequence of (ϱ_k). We have just shown that $\varrho_k \leq \mathbb{D}(Z_0, \mathcal{S}) < \infty$. Moreover $\varrho_k \geq 0$. Using the definition of the deviation, we find that there exists for each $k \in \mathbb{N}$, $\tilde{z}^k \in \mathcal{S}_k^{\varepsilon_k}$ such that $\varrho_k \leq \mathrm{dist}(\tilde{z}^k, \mathcal{S}) + 1/k$. Corollary 1 ensures that $(\tilde{z}^k)_K$ has a further subsequence $(\tilde{z}^k)_{K_1}$ that strongly converges to some $\bar{z} \in \mathcal{S}$. Since $\mathcal{S}$ is nonempty, $\mathrm{dist}(\cdot, \mathcal{S})$ is (Lipschitz) continuous [1, Thm. 3.16]. It follows that $\mathrm{dist}(\tilde{z}^k, \mathcal{S}) \to 0$ as $K_1 \ni k \to \infty$. Hence $\varrho_k \to 0$ as $K_1 \ni k \to \infty$. Since each subsequence of (ϱ_k) has a further subsequence converging to zero, $\varrho_k \to 0$ as $k \to \infty$.

Proposition 2 demonstrates that epiconvergence can imply Mosco-epiconvergence. This result is particularly relevant for PDE-constrained problems in which operators of the type $\mathbf{B}$ often appear. If V is a Banach space and Y is a complete metric space, we refer to a mapping $G : V \to Y$ as *completely continuous* if $(v^k) \subset V$ and $v^k \rightharpoonup v \in V$ implies $G(v^k) \to G(v)$.

PROPOSITION 2. *Let $Z_0 \subset Z$ be a closed, convex subset of a reflexive Banach space Z and let $Y_0 \subset Y$ be a closed subset of a Banach space Y. Suppose that $\mathbf{B} : Z \to Y$ is linear and completely continuous with $\mathbf{B}(Z_0) \subset Y_0$. If $h_k : Y_0 \to (-\infty, \infty]$ epiconverges to $h : Y_0 \to (-\infty, \infty]$ and $h_k \circ \mathbf{B} : Z_0 \to (-\infty, \infty]$ epiconverges to $h \circ \mathbf{B} : Z_0 \to (-\infty, \infty]$, then $h_k \circ \mathbf{B} : Z_0 \to (-\infty, \infty]$ Mosco-epiconverges to $h \circ \mathbf{B} : Z_0 \to (-\infty, \infty]$.*

Proof Fix $\bar{z} \in Z_0$. Let $(z^k) \subset Z_0$ be a sequence with $z^k \rightharpoonup \bar{z}$. We have $\bar{z} \in Z_0$ and $\mathbf{B}z^k$, $\mathbf{B}\bar{z} \in Y_0$. The complete continuity of $\mathbf{B}$ yields $\mathbf{B}z^k \to \mathbf{B}\bar{z}$ as $k \to \infty$. Since (h_k) epiconverges to h, $\liminf_{k \to \infty} h_k(\mathbf{B}z^k) \geq h(\mathbf{B}\bar{z})$. The hypotheses ensure that $h_k \circ \mathbf{B}$ epiconverges to $h \circ \mathbf{B}$. Putting together the pieces, we conclude that $(h_k \circ \mathbf{B})$ Mosco-epiconverges to $h \circ \mathbf{B}$.

3. Consistency of Empirical Approximations We consider the potentially infinite-dimensional risk-averse stochastic program

$$\min_{z \in Z_{\mathrm{ad}}} \mathcal{R}[F(\mathbf{B}z)] + \wp(z), \tag{9}$$

where

$$F(y)(\omega) := f(y, \xi(\omega)).$$

Let $\xi : \Omega \to \Xi$ be a random element and $\xi^1, \xi^2, \ldots$ defined on $(\Omega, \mathcal{F}, P)$ be independent identically distributed Ξ-valued random elements with law $\mathbb{P} = P \circ \xi^{-1}$. Let $\mathcal{R} : L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ be a law invariant risk measure. For $y \in Y$ and $N \in \mathbb{N}$, let $\hat{H}_{y,N} = \hat{H}_{y,N,\omega}$ be the empirical distribution function of the sample $f(y, \xi^1), \ldots, f(y, \xi^N)$. The empirical approximation of (9) is given by

$$\min_{z \in Z_{\mathrm{ad}}} \mathcal{R}[\hat{H}_{\mathbf{B}z,N}] + \wp(z). \tag{10}$$

Recall from our discussion on law invariant risk measures that $\mathcal{R}[\hat{H}_{\mathbf{B}z,N}]$ means the risk measure $\mathcal{R}$ does not distinguish between z, N-dependent random variables with distribution functions equivalent to $\hat{H}_{\mathbf{B}z,N}$.

ASSUMPTION 1.   1. The space Z is a separable, reflexive Banach space, $Z_{\mathrm{ad}} \subset Z$ is nonempty, closed, convex and bounded. The space Y is a Banach space, and $Y_0 \subset Y$ is closed, and $p \in [1, \infty)$.
2. The mapping $\mathbf{B} : Z \to Y$ is linear and completely continuous and $\mathbf{B}(Z_{\mathrm{ad}}) \subset Y_0$.
3. The function $\wp : Z \to [0, \infty)$ is convex and continuous.
4. The function $f : Y_0 \times \Xi \to \mathbb{R}$ is a Carathéodory function.
5. For all $y \in Y_0$, $f(y, \xi(\cdot)) \in L^p(\Omega, \mathcal{F}, P)$ and $F : Y_0 \to L^p(\Omega, \mathcal{F}, P)$ is continuous.
6. For each $\bar{y} \in Y_0$, there exists a neighborhood $\mathcal{Y}_{\bar{y}}$ of $\bar{y}$ and a random variable $h \in L^p(\Omega, \mathcal{F}, P)$ such that $f(y, \xi(\cdot)) \geq h(\cdot)$ for all $y \in \mathcal{V}_{\bar{y}}$.

Let m^* be the optimal value of problem (9) and let $\mathscr{S}$ be its solution set. Let $r \geq 0$. Furthermore, let $\hat{m}_N^*$ be the optimal value of (10) and let $\hat{\mathscr{S}}_N^r$ be its set of r-minimizers.

THEOREM 2.   *Let Assumption 1 hold. Suppose further that $(\Omega, \mathcal{F}, P)$ is nonatomic and complete. Let $\mathcal{R} : L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ be a convex, law invariant risk measure. If $(r_N) \subset [0, \infty)$ is a deterministic sequence such that $r_N \to 0$ as $N \to \infty$, then $\hat{m}_N^* \to m^*$ w.p. 1 as $N \to \infty$. If furthermore $\wp$ is an R-function, then $\mathbb{D}(\hat{\mathscr{S}}_N^{r_N}, \mathscr{S}) \to 0$ w.p. 1 as $N \to \infty$.*

To establish Theorem 2, we verify the hypotheses of Corollaries 2 and 4 (in the Appendix).

LEMMA 2.   *If Assumption 1 holds and $\mathcal{R} : L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a convex risk measure, then $Z \ni z \mapsto \mathcal{R}[F(\mathbf{B}z)]$ is completely continuous.*

Proof Since $\mathcal{R}$ is a finite-valued convex risk measure, it is continuous [56, Cor. 3.1]. Assumption 1 ensures the continuity of $y \mapsto F(y)$. Hence $y \mapsto \mathcal{R}[F(y)]$ is continuous. Now the complete continuity of $\mathbf{B}$ implies that of $z \mapsto \mathcal{R}[F(\mathbf{B}z)]$.

LEMMA 3.   *Let Assumption 1 hold. Suppose further that $(\Omega, \mathcal{F}, P)$ is nonatomic and complete. Let $\mathcal{R} : L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ be a convex, law invariant risk measure. Then $Y \times \Omega \ni (y, \omega) \mapsto \mathcal{R}[\hat{H}_{y,N,\omega}]$ is a Carathéodory function.*

Proof Let $([0,1], \Sigma, \nu)$ be a probability space with Σ the Lebesgue sigma-field of $[0,1]$ and ν the uniform distribution on $[0,1]$. We recall $\hat{H}_{y,N,\omega}(t) = (1/N) \sum_{i=1}^N \mathbf{1}_{(-\infty,t]}(f(y, \xi^i(\omega)))$ for $t \in \mathbb{R}$. Let $f(y, \xi^{(1)}) \leq \cdots \leq f(y, \xi^{(N)})$ be the order statistics of the sample $f(y, \xi^1), \ldots, f(y, \xi^N)$. For $q \in (0, 1]$ and $y \in Y$, we have $\hat{H}_{y,N,\omega}^{-1}(q) = f(y, \xi^{(j)}(\omega))$ if $q \in ((j-1)/N, j/N]$ irrespective of whether the sample is distinct.

We show that $y \mapsto \mathcal{R}[\hat{H}_{y,N,\omega}]$ is continuous for each $\omega \in \Omega$. Let $y^k \to y$ and fix $\omega \in \Omega$. Using the fact that ν is the uniform distribution, we have

$$\int_{[0,1]} |\hat{H}_{y^k,N,\omega}^{-1}(q) - \hat{H}_{y,N,\omega}^{-1}(q)|^p \mathrm{d}\nu(q) = \frac{1}{N} \sum_{i=1}^N |f(y^k, \xi^i(\omega)) - f(y, \xi^i(\omega))|^p.$$

Since f is a Carathéodory function and $p \in [1, \infty)$, it follows that $\hat{H}_{y^k,N,\omega}^{-1} \to \hat{H}_{y,N,\omega}^{-1}$ in $L^p([0,1], \Sigma, \nu)$. Since $(\Omega, \mathcal{F}, P)$ is nonatomic and complete, there exists a random variable $Q : \Omega \to [0,1]$ with law ν, that is, $P \circ Q^{-1} = \nu$ [19, Prop. A.5]. The functions $X_{\omega,k} : \Omega \to \mathbb{R}$ defined by $X_{\omega,k} := \hat{H}_{y^k,N,\omega}^{-1}(Q)$ are random variables with distribution function $\hat{H}_{y^k,N,\omega}$ [18, Prop. 9.1.2]. Similarly, $X_\omega := \hat{H}_{y,N,\omega}^{-1}(Q)$ is a random variable with distribution function $\hat{H}_{y,N,\omega}$ [18, Prop. 9.1.2]. Combined with the law invariance of $\mathcal{R}$, we have $\mathcal{R}[X_{\omega,k}] = \mathcal{R}[\hat{H}_{y^k,N,\omega}]$ and $\mathcal{R}[X_\omega] = \mathcal{R}[\hat{H}_{y,N,\omega}]$. Since $\hat{H}_{y^k,N,\omega}^{-1} \to \hat{H}_{y,N,\omega}^{-1}$ in $L^p([0,1], \Sigma, \nu)$, a change of variables (see, e.g., Theorem 3.6.1 in [11]) yields $X_{\omega,k} \to X_\omega$ in $L^p(\Omega, \mathcal{F}, P)$. Indeed, using $P \circ Q^{-1} = \nu$, we have

$$\int_\Omega |X_{\omega,k}(\bar{\omega}) - X_\omega(\bar{\omega})|^p \mathrm{d}P(\bar{\omega}) = \int_\Omega |\hat{H}_{y^k,N,\omega}^{-1}(Q(\bar{\omega})) - \hat{H}_{y,N,\omega}^{-1}(Q(\bar{\omega}))|^p \mathrm{d}P(\bar{\omega})$$
$$= \int_{[0,1]} |\hat{H}_{y^k,N,\omega}^{-1}(q) - \hat{H}_{y,N,\omega}^{-1}(q)|^p \mathrm{d}\nu(q).$$

Combined with the continuity of $\mathcal{R}$ [56, Cor. 3.1], we have $\mathcal{R}[X_{\omega,k}] \to \mathcal{R}[X_\omega]$. Consequently, $y \mapsto \mathcal{R}[\hat{H}_{y,N,\omega}]$ is continuous for each $\omega \in \Omega$.

For each fixed $y \in Y$, the function $\omega \mapsto \hat{H}_{y,N,\omega}^{-1}(Q) \in L^p(\Omega, \mathcal{F}, P)$ is measurable because it is the composition of a piecewise constant and measurable functions.

Combining these arguments, we find that $(y,\omega) \mapsto \mathcal{R}[\hat{H}_{y,N,\omega}]$ is a Carathéodory mapping.

COROLLARY 3.   *Under the hypotheses of Lemma 3,* (a) $\mathscr{S}$ *is nonempty and closed,* (b) $\hat{\mathscr{S}}_N^r$ *has nonempty, closed images for each* $r \in [0,\infty)$, *and* (c) $\hat{m}_N^*$ *and* $\hat{\mathscr{S}}_N^r$ *are measurable for each* $r \in [0,\infty)$.

Proof   (a) Since the set Z_{ad} is nonempty, closed, convex, and bounded, Lemma 2 when combined with the direct method of the calculus of variations ensures the assertions.

(b) Using the properties of Z_{ad} listed in part (a), Lemma 3 when combined with the direct method of the calculus of variations and the complete continuity of $\mathbf{B}$ ensures the assertions.

(c) Since $\mathbf{B}$ is completely continuous and Z is a Banach space, $\mathbf{B}$ is continuous. Lemma 3, the continuity of $\mathbf{B}$, and Theorem 8.2.11 in [5] imply the measurability assertions.

Proof of Theorem 2 To establish the consistency statements, we verify the hypotheses of Corollaries 2 and 4. Corollary 3 ensures that $\mathscr{S}$ is nonempty. Hence $\text{dist}(\cdot, \mathscr{S})$ is (Lipschitz) continuous [1, Thm. 3.16]. Corollary 3 implies that $\hat{m}_N^*$ is measurable and that $\hat{\mathscr{S}}_N^{r_N}$ is measurable with closed, nonempty images. Combined with Theorem 8.2.11 in [5], it follows that $\mathbb{D}(\hat{\mathscr{S}}_N^{r_N}, \mathscr{S})$ is measurable.

Corollary 4 ensures that $Z_{\text{ad}} \ni z \mapsto \mathcal{R}[\hat{H}_{\mathbf{B}z,N}]$ Mosco-epiconverges to $Z_{\text{ad}} \ni z \mapsto \mathcal{R}[F(\mathbf{B}z)]$ w.p. 1 as $N \to \infty$. We have $\hat{\mathscr{S}}_N^{r_N} \subset Z_{\text{ad}}$. Moreover, $\hat{\mathscr{S}}_N^{r_N}$ and $\mathscr{S} \subset Z_{\text{ad}}$ are nonempty, and $\wp$ is continuous and convex. Now, Corollary 2 ensures that w.p. 1, $\hat{m}_N^* \to m^*$ as $N \to \infty$. If furthermore $\wp$ is an R-function, then Corollary 2 ensures w.p. 1, $\mathbb{D}(\hat{\mathscr{S}}_N^{r_N}, \mathscr{S}) \to 0$ as $N \to \infty$. Since $\hat{m}_N^*$ and $\mathbb{D}(\hat{\mathscr{S}}_N^{r_N}, \mathscr{S})$ are measurable, we obtain the almost sure convergence statements.

4. Applications   We conclude with the application of our main result, Theorem 2, to several problem classes.

4.1. Consistency of Epi-Regularized and Smoothed Empirical Approximations Using Theorem 2, we demonstrate the consistency of solutions to epi-regularized and smoothed risk-averse programs using the average value-at-risk. These types of risk measures are popular in numerical approaches, see [34, 36, 37, 6, 67, 15]. For $\beta \in [0,1)$, the average value-at-risk $\text{AVaR}_\beta : L^2(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is defined by

$$\text{AVaR}_\beta[X] = \inf_{t \in \mathbb{R}} \{ t + \tfrac{1}{1-\beta} \mathbb{E}[(X-t)^+] \},$$

where $(x)^+ = \max\{0, x\}$ for $x \in \mathbb{R}$. Throughout the section, m^* and $\mathscr{S}$ denotes the optimal value and solution set of (9), respectively, with the risk measure $\mathcal{R} = \text{AVaR}_\beta$. Moreover, we denote by m_N^* and $\hat{\mathscr{S}}_N^{r_N}$ their empirical counterparts. The average value-at-risk AVaR_β is a law invariant risk measure [62].

Epi-regularization of risk measures has been proposed and analyzed in [36]. We apply the epi-regularization to the average value-at-risk. We define $\Phi : L^2(\Omega, \mathcal{F}, P) \to (-\infty, \infty]$ by

$$\Phi[X] := (1/2)\mathbb{E}[X^2] + \mathbb{E}[X].$$

For $\varepsilon > 0$, the epi-regularization $\text{AVaR}_\beta^\varepsilon : L^2(\Omega, \mathcal{F}, P) \to \mathbb{R}$ of AVaR_β is given by

$$\text{AVaR}_\beta^\varepsilon[X] := \inf_{Y \in L^2(\Omega, \mathcal{F}, P)} \{ \text{AVaR}_\beta[X-Y] + \varepsilon \Phi[\varepsilon^{-1}Y] \}. \tag{11}$$

The risk functional $\text{AVaR}_\beta^\varepsilon$ can be shown to be law invariant. See Appendix B.

For $\varepsilon > 0$, we consider the epi-regularized empirical average value-at-risk optimization problem

$$\min_{z \in Z_{\mathrm{ad}}} \{ \mathrm{AVaR}_{\beta}^{\varepsilon}[\hat{H}_{\mathbf{B}z,N}] + \wp(z) \}.$$

We let $\hat{m}_{\mathrm{epi},N}^{\varepsilon}$ be its optimal value and $\hat{\mathscr{S}}_{\mathrm{epi},N}^{\varepsilon}$ be its solution set. Note that for fixed $\varepsilon > 0$, our main result, Theorem 2, already provides an asymptotic consistency result. However, in numerical procedures, the ε-parameter is typically driven to zero. Therefore, we prove a stronger statement here.

PROPOSITION 3. *Let Assumption 1 hold. Suppose further that $(\Omega, \mathcal{F}, P)$ is nonatomic and complete. Let $(\varepsilon_N) \subset (0, \infty)$ with $\varepsilon_N \to 0$ as $N \to \infty$. Then $\hat{m}_{\mathrm{epi},N}^{\varepsilon_N} \to m^*$ w.p. 1 as $N \to \infty$. If furthermore $\wp$ is an R-function, then $\mathbb{D}(\hat{\mathscr{S}}_{\mathrm{epi},N}^{\varepsilon_N}, \mathscr{S}) \to 0$ w.p. 1 as $N \to \infty$.*

The proof of Proposition 3 is based on the following result.

LEMMA 4. *Fix $\varepsilon > 0$. The functional $\mathrm{AVaR}_{\beta}^{\varepsilon} : L^2(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a law invariant, convex risk measure. For all $X \in L^2(\Omega, \mathcal{F}, P)$, it holds that*

$$\mathrm{AVaR}_{\beta}[X] - \tfrac{\varepsilon \beta}{2(1-\beta)} \leq \mathrm{AVaR}_{\beta}^{\varepsilon}[X] \leq \mathrm{AVaR}_{\beta}[X].$$

Proof The functional $\mathrm{AVaR}_{\beta}^{\varepsilon}$ is a convex risk measure [36, pp. 3 and 5–6]. By the arguments in Appendix B, it is law invariant. Since $\Phi[0] = 0$, the second estimate is implied by Proposition 1 in [36]. Fix $X \in L^2(\Omega, \mathcal{F}, P)$. Since AVaR_{β} is subdifferentiable [63, p. 243], Proposition 2 in [36] yields for all $\vartheta \in \partial \mathrm{AVaR}_{\beta}[X]$,

$$\mathrm{AVaR}_{\beta}^{\varepsilon}[X] \geq \mathrm{AVaR}_{\beta}[X] - \varepsilon \Phi^*[\vartheta].$$

Let $\vartheta \in \partial \mathrm{AVaR}_{\beta}[X]$ be arbitrary. We have $0 \leq \vartheta \leq 1/(1 - \beta)$ w.p. 1, $\mathbb{E}[\vartheta] = 1$ [63, p. 243] and $\Phi^*[\vartheta] = (1/2)\mathbb{E}[(\vartheta - 1)^2]$; see Remark 5 in [36]. Here Φ^* is the Fenchel conjugate to Φ. Hence

$$\Phi^*[\vartheta] = (1/2)\mathbb{E}[\vartheta^2] - \mathbb{E}[\vartheta] + (1/2) = (1/2)\mathbb{E}[\vartheta^2] - (1/2) \leq \frac{1}{2} \frac{1 - (1 - \beta)}{1 - \beta} = \frac{1}{2} \frac{\beta}{1 - \beta}.$$

Proof of Proposition 3 Following the proof of Corollary 3 and using the fact that $\mathrm{AVaR}_{\beta}^{\varepsilon_N}$ is a law invariant, convex risk measure (see Lemma 4), we find that $\hat{m}_{\mathrm{epi},N}^{\varepsilon_N}$ and $\mathscr{S}_{\mathrm{epi},N}^{\varepsilon_N}$ are measurable. Lemma 4 ensures that $\hat{m}_N^* - \tfrac{\varepsilon_N \beta}{2(1-\beta)} \leq \hat{m}_{\mathrm{epi},N}^{\varepsilon_N} \leq \hat{m}_N^*$. Applying Theorem 2 with $\mathcal{R} = \mathrm{AVaR}_{\beta}$ yields $\hat{m}_N^* \to m^*$ w.p. 1 as $N \to \infty$. Combined with $\varepsilon_N \to 0$, we find that $\hat{m}_{\mathrm{epi},N}^{\varepsilon_N} \to m^*$ w.p. 1 as $N \to \infty$.

If $z_N^{\varepsilon_N} \in \hat{\mathscr{S}}_{\mathrm{epi},N}^{\varepsilon_N}$, then Lemma 4 ensures that $z_N^{\varepsilon_N} \in \hat{\mathscr{S}}_N^{r_N}$, where $r_N \coloneqq \tfrac{\varepsilon_N \beta}{2(1-\beta)}$. Hence $\hat{\mathscr{S}}_{\mathrm{epi},N}^{\varepsilon_N} \subset \hat{\mathscr{S}}_N^{r_N}$. Using (4), we find that $\mathbb{D}(\hat{\mathscr{S}}_{\mathrm{epi},N}^{\varepsilon_N}, \mathscr{S}) \leq \mathbb{D}(\hat{\mathscr{S}}_N^{r_N}, \mathscr{S})$. Applying Theorem 2 with $\mathcal{R} = \mathrm{AVaR}_{\beta}$ yields the second assertion.

Next, we establish the consistency of solutions to smoothed average value-at-risk problems using a smoothing function for $(\cdot)^+$. For brevity, we focus on a particular smoothing function for the plus function $(\cdot)^+$. For $\varepsilon > 0$, we define the smoothed plus function $(\cdot)_{\varepsilon}^+ : \mathbb{R} \to \mathbb{R}$ by

$$(x)_{\varepsilon}^+ \coloneqq \varepsilon \ln(1 + \exp(x/\varepsilon)).$$

Using $(\cdot)_{\varepsilon}^+$, we define the smoothed average value-at-risk $\sigma_{\beta}^{\varepsilon} : L^2(\Omega, \mathcal{F}, P) \to \mathbb{R}$ by

$$\sigma_{\beta}^{\varepsilon}[X] \coloneqq \inf_{t \in \mathbb{R}} \{ t + \tfrac{1}{1-\beta} \mathbb{E}[(X - t)_{\varepsilon}^+] \}. \tag{12}$$

This version of the smoothed average value-at-risk has been used in [67] for stochastic stellarator coil design and in [6] for adaptive sampling techniques for risk-averse optimization. See the Appendix B for a short proof of its law invariance.

For $\varepsilon > 0$, we consider the smoothed empirical average value-at-risk optimization problem

$$\min_{z \in Z_{\text{ad}}} \{ \sigma_\beta^\varepsilon[\hat{H}_{\mathbf{B}z,N}] + \wp(z) \},$$

We let $\hat{m}_{\text{s},N}^\varepsilon$ be its optimal value and $\hat{\mathscr{S}}_{\text{s},N}^\varepsilon$ be its solution set.

PROPOSITION 4. *Let Assumption 1 hold. Suppose further that $(\Omega, \mathcal{F}, P)$ is nonatomic and complete. Let $(\varepsilon_N) \subset (0, \infty)$ with $\varepsilon_N \to 0$ as $N \to \infty$. Then $\hat{m}_{\text{s},N}^{\varepsilon_N} \to m^*$ w.p. 1 as $N \to \infty$. If furthermore $\wp$ is an R-function, then $\mathbb{D}(\hat{\mathscr{S}}_{\text{s},N}^{\varepsilon_N}, \mathscr{S}) \to 0$ w.p. 1 as $N \to \infty$.*

Proposition 4 is established using Lemma 5.

LEMMA 5. *Fix $\varepsilon > 0$. The functional $\sigma_\beta^\varepsilon \colon L^2(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a law invariant, convex risk measure. For all $X \in L^2(\Omega, \mathcal{F}, P)$, it holds that*

$$\text{AVaR}_\beta[X] \leq \sigma_\beta^\varepsilon[X] \leq \text{AVaR}_\beta[X] + \ln(2)\varepsilon/(1-\beta).$$

Proof The smoothed average value-at-risk σ_β^ε is a convex risk measure [34, Props. 4.4–4.6]. By the arguments in Appendix B, it is law invariant. For $x \in \mathbb{R}$, we have $(x)^+ \leq (x)_\varepsilon^+ \leq (x)^+ + \varepsilon \ln(2)$, yielding the error bounds.

Proof of Proposition 4 The proof is similar to that of Proposition 3. Following the proof of Corollary 3 and using the fact that $\sigma_\beta^{\varepsilon_N}$ is a law invariant, convex risk measure (see Lemma 5), we find that $\hat{m}_{\text{s},N}^{\varepsilon_N}$ and $\hat{\mathscr{S}}_{\text{s},N}^{\varepsilon_N}$ are measurable. Lemma 4 ensures that $\hat{m}_N^* \leq \hat{m}_{\text{s},N}^{\varepsilon_N} \leq \hat{m}_N^* + \ln(2)\varepsilon_N/(1-\beta)$. Applying Theorem 2 with $\mathcal{R} = \text{AVaR}_\beta$ yields $\hat{m}_N^* \to m^*$ w.p. 1 as $N \to \infty$. Combined with $\varepsilon_N \to 0$, we find that $\hat{m}_{\text{s},N}^{\varepsilon_N} \to m^*$ w.p. 1 as $N \to \infty$.

If $z_N^{\varepsilon_N} \in \hat{\mathscr{S}}_{\text{s},N}^{\varepsilon_N}$, then Lemma 4 ensures that $z_N^{\varepsilon_N} \in \hat{\mathscr{S}}_N^{r_N}$, where $r_N := \ln(2)\varepsilon_N/(1-\beta)$. Hence $\hat{\mathscr{S}}_{\text{s},N}^{\varepsilon_N} \subset \hat{\mathscr{S}}_N^{r_N}$. Using (4), we find that $\mathbb{D}(\hat{\mathscr{S}}_{\text{s},N}^{\varepsilon_N}, \mathscr{S}) \leq \mathbb{D}(\hat{\mathscr{S}}_N^{r_N}, \mathscr{S})$. Applying Theorem 2 with $\mathcal{R} = \text{AVaR}_\beta$ yields the second assertion.

4.2. Risk-Averse Semilinear PDE-Constrained Optimization Our consistency result, Theorem 2, is applicable to risk-averse semilinear PDE-constrained optimization as we demonstrate in this section. Following [37] (see also [23, 24]), we consider

$$\min_{z \in Z_{\text{ad}}} (1/2)\mathcal{R}[\|(1 - \iota S(z))^+\|_{L^2(D)}^2] + (\alpha/2)\|z\|_{L^2(D)}^2, \tag{13}$$

where $\alpha > 0$, $\iota \colon H^1(D) \to L^2(D)$ is the embedding operator of the compact embedding $H^1(D) \hookrightarrow L^2(D)$, $Z_{\text{ad}} := \{ z \in L^2(D) \colon \mathfrak{l} \leq z \leq \mathfrak{u} \}$ with $\mathfrak{l}, \mathfrak{u} \in L^2(D)$ and $\mathfrak{l} \leq \mathfrak{u}$, and for each $(z, \xi) \in L^2(D) \times \Xi$, $S(z)(\xi) \in H^1(D)$ is the solution to:

$$\text{find} \quad u \in H^1(D) \colon \quad \mathbf{A}(u, \xi) = \mathbf{B}_1(\xi)\iota^* z + \mathbf{b}(\xi), \tag{14}$$

where $\mathbf{A} \colon H^1(D) \times \Xi \to H^1(D)^*$, $\mathbf{B}_1 \colon \Xi \to \mathscr{L}(H^1(D)^*, H^1(D)^*)$, and $\mathbf{b} \colon \Xi \to H^1(D)^*$ are defined by

$$\langle \mathbf{A}(u, \xi), v \rangle_{H^1(D)^*, H^1(D)} := \int_D a(\xi)(x)[\nabla u(x)^T \nabla v(x) + u(x)v(x)]\mathrm{d}x + \int_D u(x)^3 v(x)\mathrm{d}x,$$

$$\langle \mathbf{B}_1(\xi)y, v \rangle_{H^1(D)^*, H^1(D)} := \int_D [B(\xi)y](x)v(x)\mathrm{d}x, \quad \langle \mathbf{b}(\xi), v \rangle_{H^1(D)^*, H^1(D)} := \int_D b(\xi)(x)v(x)\mathrm{d}x.$$

Here, $b \colon \Xi \to L^2(D)$ is essentially bounded, $a \colon \Xi \to C^0(\bar{D})$ is measurable and there exists constants $\kappa_{\min}, \kappa_{\max} > 0$ such that $\kappa_{\min} \leq a(\xi)(x) \leq \kappa_{\max}$ for all $(\xi, x) \times \Xi \times \bar{D}$. It remains to define

$B(\xi)\colon H^1(D)^* \to H^1(D)$. Fix $(y, \xi) \in H^1(D)^* \times \Xi$. We define $B(\xi)y \in H^1(D)$ as the solution to: find $w \in H^1(D)$ such that

$$\int_D [r(\xi)\nabla w(x)^T v(\xi) + w(x)v(x)]\mathrm{d}x = \langle y, v \rangle_{H^1(D)^*, H^1(D)} \quad \text{for all} \quad v \in H^1(D),$$

where $r : \Xi \to (0, \infty)$ is random variable such that there exists $r_{\min}$, $r_{\max} > 0$ with $r_{\min} \leq r(\xi) \leq r_{\max}$ for all $\xi \in \Xi$. Since $\iota u = u$ for all $u \in H^1(D)$, we have $\langle \iota^* z, v \rangle_{H^1(D)^*, H^1(D)} = (z, v)_{L^2(D)}$ for all $z \in L^2(D)$ and $v \in H^1(D)$ [12, p. 21].

We express (13) in the form given in (9) and verify Assumption 1. For each $(y, \xi) \in H^1(D)^* \times \Xi$, we consider the auxiliary random operator equation

$$\text{Find} \quad u \in H^1(D)\colon \quad \mathbf{A}(u, \xi) = \mathbf{B}_1(\xi)y + \mathbf{b}(\xi). \tag{15}$$

LEMMA 6. *Under the above hypotheses, for each $(y, \xi) \in H^1(D)^* \times \Xi$, the operator equation (15) has a unique solution $\widetilde{S}(y)(\xi)$, $\widetilde{S}(y) \in L^q(\Xi, \mathcal{A}, \mathbb{P}; H^1(D))$ for each $q \in [1, \infty]$ and $y \in H^1(D)^*$, and $\widetilde{S} : H^1(D)^* \to L^q(\Xi, \mathcal{A}, \mathbb{P}; H^1(D))$ is Lipschitz continuous for each $q \in [1, \infty]$.*

Let U be a reflexive Banach space. We recall that an operator $A : U \to U^*$ is κ-strongly monotone if there exists $\kappa > 0$ such that

$$\langle A(u_2) - A(u_1), u_2 - u_1 \rangle_{U^*, U} \geq \kappa \|u_2 - u_1\|_U^2 \quad \text{for all} \quad u_1, u_2 \in U.$$

Proof of Lemma 6 For each $\xi \in \Xi$, $\mathbf{A}(\cdot, \xi)$ is $\kappa_{\min}$-strongly monotone and it holds that

$$\|\mathbf{B}_1(\xi)\|_{\mathscr{L}(H^1(D)^*, H^1(D))} \leq 1/\min\{r_{\min}, 1\};$$

cf. [37, p. 13]. The existence, uniqueness and the stability estimate

$$\|\widetilde{S}(y)(\xi)\|_{H^1(D)} \leq (1/\kappa_{\min})\|\mathbf{B}_1(\xi)y\|_{H^1(D)^*} + (1/\kappa_{\min})\|\mathbf{b}(\xi)\|_{H^1(D)^*}$$

are a consequence of the Minty–Browder theorem [68, Thm. A.26], for example. Using Filippov's theorem [5, Thm. 8.2.10], we can show that $\widetilde{S}(y)$ is measurable. Combined with the stability estimate and Hölder's inequality, we conclude that $\widetilde{S}(y) \in L^q(\Xi, \mathcal{A}, \mathbb{P}; H^1(D))$ for each $q \in [1, \infty]$ and $y \in H^1(D)^*$. Since for all $y_1, y_2 \in H^1(D)^*$ and $\xi \in \Xi$, we have (cf. [37, eq. (3.7)])

$$\|\widetilde{S}(y_2)(\xi) - \widetilde{S}(y_1)(\xi)\|_U \leq (1/\kappa_{\min})\|\mathbf{B}_1(\xi)[y_2 - y_1]\|_{H^1(D)^*},$$

the mapping $\widetilde{S} : H^1(D)^* \to L^q(\Xi, \mathcal{A}, \mathbb{P}; H^1(D))$ is Lipschitz continuous for all $q \in [1, \infty]$.

The function $\wp$ defined by $\wp(z) := (\alpha/2)\|z\|_{L^2(D)}^2$ is an R-function according to Lemma 1, as $\alpha > 0$ and $L^2(D)$ is a Hilbert space and hence has the Radon–Riesz property [12, Prop. 2.35]. The operator $\mathbf{B} := \iota^*$ with ι^* being the adjoint operator to ι is linear and completely continuous because ι is a compact operator by the Sobolev embedding theorem. We define $f : H^1(D)^* \times \Xi \to [0, \infty)$ by $f(y, \xi) := (1/2)\|(1 - \iota \widetilde{S}(y)(\xi))^+\|_{L^2(D)}^2$. The mapping $\mathcal{J} : L^2(\Xi, \mathcal{A}, \mathbb{P}; H^1(D)) \to [0, \infty)$ given by $\mathcal{J}(y) = (1/2)\|(1 - \iota y)^+\|_{L^2(D)}^2$ is continuous [37, p. 14]. Since $F = \mathcal{J} \circ \widetilde{S}$, Lemma 6 ensures that $F : H^1(D)^* \to L^2(\Xi, \mathcal{A}, \mathbb{P}; H^1(D))$ is well-defined and continuous. Having verified Assumption 1, we can apply Theorem 2 to study the consistency of empirical approximations of (13).

4.3. Risk-Averse Optimization with Variational Inequalities   We consider a risk-averse optimization problem governed by an elliptic variational inequality with random inputs. Our presentation is inspired by that in [26]. We consider

$$\min_{z \in Z_{\mathrm{ad}}} (1/2)\mathcal{R}[\|\iota S(z) - u_d\|_{L^2(D)}^2] + (\alpha/2)\|z\|_{L^2(D)}^2, \tag{16}$$

where $\alpha > 0$, $u_d \in L^2(D)$, $\iota : H_0^1(D) \to L^2(D)$ is the embedding operator of the compact embedding $H^1(D) \hookrightarrow L^2(D)$, and Z_{ad} is as in Section 4.2. For each $(z, \xi) \in L^2(D) \times \Xi$, $S(z)(\xi) \in H_0^1(D)$ is the solution to the parameterized elliptic variational inequality:

$$\text{find} \quad u \in K_\psi : \quad \langle A(\xi)u - \iota^* z, v - u \rangle_{H^{-1}(D), H_0^1(D)} \geq 0 \quad \text{for all} \quad v \in K_\psi, \tag{17}$$

where ι^* is the adjoint operator to ι, $H^{-1}(D) := H_0^1(D)^*$, $A : \Xi \to \mathscr{L}(H_0^1(D), H^{-1}(D))$ is a parameterized elliptic operator, and $K_\psi := \{ u \in H_0^1(D) : u \geq \psi \}$ with $\psi \in H^1(D)$ and $\psi_{\partial D} \leq 0$ is the obstacle. The set K_ψ is nonempty [66, p. 129]. For $(y, \xi) \in H^{-1}(D) \times \Xi$, we also consider the auxiliary parameterized elliptic variational inequality:

$$\text{find} \quad u \in K_\psi : \quad \langle A(\xi)u - y, v - u \rangle_{H^{-1}(D), H_0^1(D)} \geq 0 \quad \text{for all} \quad v \in K_\psi. \tag{18}$$

If $\widetilde{S}(y)(\xi)$ with $y = \iota^* z$ is a solution to (18), then it is a solution to (17).

We assume that $A : \Xi \to \mathscr{L}(H_0^1(D), H^{-1}(D))$ is uniformly measurable, that is, there exists a sequence $A_k : \Xi \to \mathscr{L}(H_0^1(D), H^{-1}(D))$ of simple mappings such that $A_k(\xi) \to A(\xi)$ in $\mathscr{L}(H_0^1(D), H^{-1}(D))$ as $k \to \infty$ for each $\xi \in \Xi$. Moreover, we assume that there exists constants $\kappa_{\min}$, $\kappa_{\max} > 0$ such that for each $\xi \in \Xi$, $A(\xi)$ is $\kappa_{\min}$-strongly monotone and $\|A(\xi)\|_{\mathscr{L}(H_0^1(D), H^{-1}(D))} \leq \kappa_{\max}$. Under these conditions, the auxiliary variational inequality (17) has a unique solution $\widetilde{S}(z)(\xi)$ for each $(y, \xi) \in H^{-1}(D) \times \Xi$ and $\widetilde{S}(\cdot)(\xi)$ is Lipschitz continuous with Lipschitz constant $1/\kappa_{\min}$; cf. [26, Thm. 7.3]. Using results established in [25, p. 180], we can show that $\widetilde{S}(y) \in L^q(\Xi, \mathcal{A}, \mathbb{P}; H_0^1(D))$ for all $q \in [1, \infty]$ and $y \in H^{-1}(D)$. Combined with the Lipschitz continuity, we find that $\widetilde{S} : H^{-1}(D) \to L^q(\Xi, \mathcal{A}, \mathbb{P}; H_0^1(D))$ is continuous for each $q \in [1, \infty]$.

We express (16) in the form given in (9) and verify Assumption 1. The function $\wp$ defined by $\wp(z) := (\alpha/2)\|z\|_{L^2(D)}^2$ is an R-function; see Section 4.2. The operator $\mathbf{B} := \iota^*$ is linear and completely continuous because ι is a compact operator. We define $f : H^{-1}(D) \times \Xi \to [0, \infty)$ by $f(y, \xi) := (1/2)\|\iota \widetilde{S}(y)(\xi) - u_d\|_{L^2(D)}^2$. The mapping $\mathcal{J} : L^2(\Xi, \mathcal{A}, \mathbb{P}; H_0^1(D)) \to [0, \infty)$ given by $\mathcal{J}(y) := (1/2)\|\iota y - u_d\|_{L^2(D)}^2$ is continuous [35, Example 3.2 and Theorem 3.5]. Since $F = \mathcal{J} \circ \widetilde{S}$, $F : H^{-1}(D) \to L^2(\Xi, \mathcal{A}, \mathbb{P}; H_0^1(D))$ is well-defined and continuous. Having verified Assumption 1, we can apply Theorem 2, which in turn yields the consistency of empirical approximations of (16).

5. Conclusion   We have seen that consistency results, in particular, norm consistency of empirical minimizers for nonconvex, risk-averse stochastic optimization problems involving infinite-dimensional decision spaces are in fact available. The central property on which the entire discussion depends is the ability to draw compactness from the structure of the objective function. As the examples illustrate, this is much more the rule rather than the exception. In fact, even in examples such as topology optimization, [8], where the decision variable enters the PDE in a nonlinear fashion, the required use of either filters or other regularization strategies, see e.g. [39, 65], also provides compactness.

There remain many open challenges. These include applications to multistage or dynamic problems, large deviation results for optimal values and solutions, and central limit theorems. In many instances, the known techniques are limited by nonsmoothness of the risk measure $\mathcal{R}$ and the infinite-dimensional decision spaces. However, the main result in this text, Theorem 2, is a first major step and an essential tool towards verifying the convergence of numerical optimization methods that make use of empirical approximations.

Appendix A: Law of Large Numbers for Risk Functionals   We generalize the epigraphical law of large numbers for law invariant risk function established in [61] to allow for random lower semicontinuous functions defined on complete, separable metric spaces. The proof provided in [61] generalizes to this more general setting with only a few notational changes needed. Nevertheless, we verify the liminf-condition of epiconvergence using ideas from the proof of Proposition 7.1 in [55]. The limsup-condition is established as in [61].

ASSUMPTION 2.   Let $(\Omega, \mathcal{F}, P)$ be nonatomic, complete probability space, and let (Θ, Σ) be a measurable space. Let $\zeta : \Omega \to \Theta$ be a random element and $\zeta^1, \zeta^2, \ldots$ defined on $(\Omega, \mathcal{F}, P)$ be independent identically distributed Θ-valued random elements each having the same distribution as that of ζ. Let (V, d_V) be a complete, separable metric space and let $1 \leq p < \infty$.

($\mathfrak{R}1$)   The function $\Psi : V \times \Theta \to \mathbb{R}$ is random lower semicontinuous.

($\mathfrak{R}2$)   For each $v \in V$, $\Psi_v(\cdot) := \Psi(v, \zeta(\cdot)) \in L^p(\Omega, \mathcal{F}, P)$.

($\mathfrak{R}3$)   For each $\bar{v} \in V$, there exists a neighborhood $\mathcal{V}_{\bar{v}}$ of $\bar{v}$ and a random variable $h \in L^p(\Omega, \mathcal{F}, P)$ such that $\Psi(v, \zeta(\cdot)) \geq h(\cdot)$ for all $v \in \mathcal{V}_{\bar{v}}$.

Theorem 3 is as Theorem 3.1 in [61] but allows for complete, separable metric spaces V instead of $\mathbb{R}^n$. Let $\rho : L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ be a law invariant risk measure. Let $v \in V$ and let $\hat{H}_{v,N} = \hat{H}_{v,N,\omega}$ be the empirical distribution function of $\Psi(v, \zeta^1), \ldots, \Psi(v, \zeta^N)$. We define $\hat{\phi}_N : V \times \Omega \to \mathbb{R}$ and $\phi : V \to \mathbb{R}$ by $\hat{\phi}_N(v) := \hat{\phi}_N(v, \omega) := \rho(\hat{H}_{v,N})$ and $\phi(v) := \rho(\Psi_v)$.

THEOREM 3.   *If Assumption 2 holds and $\rho : L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a law invariant, convex risk measure, then ϕ is lower semicontinuous and finite-valued, and $\hat{\phi}_N$ epiconverges to ϕ w.p. 1 as $N \to \infty$.*

Before establishing Theorem 3, we formulate a law of large numbers with respect to Mosco-epiconvergence.

COROLLARY 4.   *Let $Y_0 \subset Y$ be a closed subset of a Banach space Y and let W_0 be a closed, convex subset of a reflexive, separable Banach space W. Let the hypotheses of Theorem 3 hold with $V = Y_0$. Suppose that $\mathbf{B} : W \to Y$ is linear and completely continuous with $\mathbf{B}(W_0) \subset Y_0$. Then $\hat{\phi}_N \circ \mathbf{B} : W_0 \to \mathbb{R}$ Mosco-epiconverges to $\phi \circ \mathbf{B} : W_0 \to \mathbb{R}$ w.p. 1 as $N \to \infty$.*

Proof Theorem 3 ensures that $\hat{\phi}_N$ epiconverges to ϕ w.p. 1 as $N \to \infty$. Since W_0 defines a complete separable metric space and $\mathbf{B}$ is continuous, Theorem 3 further ensures that $\hat{\phi}_N \circ \mathbf{B}$ epiconverges to $\phi \circ \mathbf{B}$ w.p. 1 as $N \to \infty$. Combined with Proposition 2 and the complete continuity of $\mathbf{B}$, we conclude that $\hat{\phi}_N \circ \mathbf{B}$ Mosco-epiconverges to $\phi \circ \mathbf{B}$ w.p. 1 as $N \to \infty$.

As already mentioned, the proof of Theorem 3 presented in [61, Thm. 3.1] for $V = \mathbb{R}^n$ can be generalized to the above setting without much effort. A key result for establishing Theorem 3 is Theorem 4. To formulate Theorem 4, let $X : \Omega \to \mathbb{R}$ be a random variable and $X_1, X_2, \ldots$ defined on $(\Omega, \mathcal{F}, P)$ be independent identically distributed real-valued random variables each having the same distribution as that of X. Moreover, let $\hat{H}_N$ be the empirical distribution function of the sample $X_1, \ldots, X_N$.

THEOREM 4 (see [**61, Thm. 2.1**] and [**63, Thm. 9.65**]).   *If $(\Omega, \mathcal{F}, P)$ is complete and nonatomic, $1 \leq p < \infty$, and $\rho : L^p(\Omega, \mathcal{F}, P) \to \mathbb{R}$ is a law invariant, convex risk measure, then $\rho(\hat{H}_N)$ converges to $\rho(X)$ w.p. 1 as $N \to \infty$.*

Proof of Theorem 3 The fact that ϕ is finite-valued and lower semicontinuous can be established as in the proof of Theorem 3.1 in [61]. To establish the epiconvergence, we make use of the constructions made in the proof of Proposition 7.1 in [55]. Proposition 7.1 in [55] establishes epiconvergence in case that $\rho(\cdot) = \mathbb{E}[\cdot]$, but without assuming $(\Omega, \mathcal{F}, P)$ be nonatomic. Let $\mathcal{E} \subset V$ be a countable

dense subset of V and Q_+ be the nonnegative rational numbers. For $v \in \mathcal{E}$ and $r \in Q_+$, we define $\pi_{v,r} : \Omega \to \mathbb{R}$ by

$$\pi_{v,r}(\omega) := \inf_{v' \in B(v,r)} \Psi(v, \zeta(\omega)) \quad \text{if} \quad r > 0 \quad \text{and} \quad \pi_{v,0}(\omega) := \Psi(v, \zeta(\omega)) \quad \text{if} \quad r = 0,$$

where $B(v,r) := \{v' \in V : d_V(v', v) < r\}$. Theorem 3.4 in [33], $(\mathfrak{R}1)$, and $(\mathfrak{R}2)$ ensure that $\pi_{v,r}$ is a real-valued random variable for each $v \in V$ and $r \in [0, \infty)$. Combined with $(\mathfrak{R}3)$, we find that for every $v \in \mathcal{E}$, there exists a neighborhood $\mathcal{V}_v$ of v and $r_v \in (0, \infty)$ such that

$$B(v, r_v) \subset \mathcal{V}_v \quad \text{and} \quad \pi_{v,r} \in L^p(\Omega, \mathcal{F}, P) \quad \text{for all} \quad r \in [0, r_v].$$

Let $\hat{\rho}_N(\pi_{v,r}, \omega)$ be the empirical estimate of $\rho(\pi_{v,r})$ based on the same sample as that used to estimate $\hat{\rho}(\cdot)$, that is, $\hat{\rho}_N(\pi_{v,r}, \omega) := \rho(\tilde{H}_{v,N,\omega})$, where $\tilde{H}_{v,N,\omega}$ is the empirical distribution function of the sample $\pi_{v,r}(\zeta^1), \ldots, \pi_{v,r}(\zeta^N)$. For every $v \in \mathcal{E}$ and $r \in [0, r_v] \cap Q_+$, Theorem 4 ensures that $\hat{\rho}_N(\pi_{v,r}) \to \rho(\pi_{v,r})$ w.p. 1 as $N \to \infty$. Since $\{(v, r) : r \in [0, r_v] \cap Q_+ : v \in \mathcal{E}\}$ is countable, there exists $\Omega_0 \subset \Omega$ with $\Omega_0 \in \mathcal{F}$ and $P(\Omega_0) = 1$ such that

$$\hat{\rho}_N(\pi_{v,r}, \omega) \to \rho(\pi_{v,r}) \quad \text{as} \quad N \to \infty \quad \text{for all} \quad \omega \in \Omega_0 \quad \text{and} \quad r \in [0, r_v] \cap Q_+, v \in \mathcal{E}.$$

Now, we verify the liminf-condition of epiconvergence. Fix $v \in V$ and fix $v_N \to v$ as $N \to \infty$. There exists $\bar{N}(\ell) \in \mathbb{N}$, $z^\ell \in \mathcal{E}$ and $r_\ell \in [0, r_v] \cap Q_+$, $\ell \in \mathbb{N}$ such that $z^\ell \to v$, $r_\ell \to 0$,

$$B(z^{\ell+1}, r_{\ell+1}) \subset B(z^\ell, r_\ell), \quad \text{and} \quad v_N \in B(z^\ell, r_\ell) \quad \text{for all} \quad N \geq \bar{N}(\ell), \ell \in \mathbb{N}.$$

Fix $\ell \in \mathbb{N}$. For all $N \geq \bar{N}(\ell)$ and $\omega \in \Omega_0$, Theorem 6.50 in [63] when combined with the fact that ρ is law invariant and monotone, and $v_N \in B(z^\ell, r_\ell)$ ensures

$$\hat{\phi}_N(v_N, \omega) \geq \hat{\rho}_N(\pi_{z^\ell, r_\ell}, \omega \hat{\phi}_N(v_N, \omega) \geq \hat{\rho}_N(\pi_{z^\ell, r_\ell}, \omega). \tag{19}$$

Moreover, for all $\omega \in \Omega_0$,

$$\hat{\rho}_N(\pi_{z^\ell, r_\ell}, \omega) \to \rho(\pi_{z^\ell, r_\ell}) \quad \text{as} \quad N \to \infty. \tag{20}$$

Since $B(z^{\ell+1}, r_{\ell+1}) \subset B(z^\ell, r_\ell)$, we have $\pi_{z^\ell, r_\ell} \leq \pi_{z^{\ell+1}, r_{\ell+1}}$. For all $\ell \in \mathbb{N}$ and for all $\omega \in \Omega$, the lower semicontinuity of $\Psi(\cdot, \zeta(\omega))$ (see $(\mathfrak{R}1)$) ensures $\pi_{z^\ell, r_\ell}(\omega) \nearrow \pi_{v,0}(\omega) = \Psi(v, \zeta(\omega))$ as $\ell \to \infty$ [33, p. 432]. Thus $\pi_{v,0} - \pi_{z^1, r_1} \geq \pi_{v,0} - \pi_{z^{\ell+1}, r_{\ell+1}} \geq 0$ for all $\ell \in \mathbb{N}$. Consequently, $|\pi_{v,0} - \pi_{z^1, r_1}|^p \geq |\pi_{v,0} - \pi_{z^{\ell+1}, r_{\ell+1}}|^p$. Since $\pi_{z^1, r_1}, \pi_{v,0} \in L^p(\Omega, \mathcal{F}, P)$, the dominated convergence theorem implies $\pi_{z^\ell, r_\ell} \to \pi_{v,0}$ as $\ell \to \infty$ in $L^p(\Omega, \mathcal{F}, P)$. Using the fact that the risk measure ρ is real-valued and convex, it follows that ρ is continuous [56, Cor. 3.1] and monotone. Consequently, $\rho(\pi_{z^\ell, r_\ell}) \nearrow \rho(\pi_{v,0}) = \phi(v)$ as $\ell \to \infty$. Combined with (19) and (20), we find that

$$\liminf_{N \to \infty} \hat{\phi}_N(v_N, \omega) \geq \phi(v).$$

Now, we verify the limsup-condition of epiconvergence using the arguments in [61]. Since ϕ is defined on a separable metric space, finite-valued and lower semicontinuous, there exists a countable set $\mathcal{D} \subset V$ such that for each $v \in V$, there exists a sequence $(v_k) \subset \mathcal{D}$ such that $v_k \to v$ and $\phi(v_k) \to \phi(v)$ as $k \to \infty$ [69, Lem. 3]. Since $\mathcal{D}$ is countable, Theorem 4 ensures the existence of $\Omega_1 \subset \Omega$ with $\Omega_1 \in \mathcal{F}$ and $P(\Omega_1) = 1$ such that for each $v \in \mathcal{D}$ and all $\omega \in \Omega_1$, we have $\hat{\phi}_N(v, \omega) \to \phi(v)$. Fix $v \in V$ and let $(v_k) \subset \mathcal{D}$ be a sequence such that $v_k \to v$ and $\phi(v_k) \to \phi(v)$ as $k \to \infty$. We now proceed with a diagonalization argument (see, e.g., Corollary 1.16 or 1.18 in [3]). For each $k \in \mathbb{N}$ and every $\omega \in \Omega_1$, we have $\hat{\phi}_N(v_k, \omega) \to \phi(v_k)$ as $N \to \infty$. Moreover $\phi(v_k) \to \phi(v)$ as $k \to \infty$. Consequently, for each $\omega \in \Omega_1$, there exists a mapping $\mathbb{N} \ni N \mapsto k_\omega(N) \in \mathbb{N}$ increasing to ∞ such that $\hat{\phi}_N(v_{k_\omega(N)}, \omega) \to \phi(v)$ as $N \to \infty$. Since $v_k \to v$ as $k \to \infty$, we further have $v_{k_\omega(N)} \to v$ as $N \to \infty$ for each $\omega \in \Omega_1$.

Appendix B: Law Invariance of $\mathrm{AVaR}_\beta^\varepsilon$ **and** σ_β^ε Both $\mathrm{AVaR}_\beta^\varepsilon$ defined in (11) and σ_β^ε given in (12) are optimized certainty equivalents in the sense of [7], i.e. they are fully characterized by convex, continuous scalar regret functions $v_{\mathrm{epi}}, v_{\mathrm{s}} : \mathbb{R} \to \mathbb{R}$ such that for each $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathrm{AVaR}_\beta^\varepsilon[X] = \inf_{t \in \mathbb{R}} \{\, t + \tfrac{1}{1-\beta} \mathbb{E}[v_{\mathrm{epi}}(X - t)] \,\},$$
$$\sigma_\beta^\varepsilon[X] = \inf_{t \in \mathbb{R}} \{\, t + \tfrac{1}{1-\beta} \mathbb{E}[v_{\mathrm{s}}(X - t)] \,\}.$$

The explicit form for v_{epi} can be found in Example 2 in [36] and $v_{\mathrm{s}}(\cdot) = (\cdot)_\varepsilon^+$ for fixed $\varepsilon > 0$.

It is not essential for the underlying probability space to be nonatomic for the law invariance of these functionals. Indeed, start by letting $v : \mathbb{R} \to \mathbb{R}$ be continuous and hence, measurable. For each $X \in L^p(\Omega, \mathcal{F}, P)$ and $t \in \mathbb{R}$, let $v(X - t)$ be integrable indendently of X and t, which is the case for both v_{epi} and v_{s}. Let $X_1, X_2 \in L^p(\Omega, \mathcal{F}, P)$ be distributionally equivalent with respect to P. Since the distribution functions of X_1 and X_2 are equal and each distribution function uniquely determines a probability law on $\mathbb{R}$ [10, Thm. 12.4], it holds that $P \circ X_1^{-1} = P \circ X_2^{-1}$. For all $t \in \mathbb{R}$, we have

$$\mathbb{E}[v(X_1 - t)] = \int_\Omega v(X_1(\omega) - t) \mathrm{d}P(\omega) = \int_\mathbb{R} v(x - t) \mathrm{d}P \circ X_1^{-1}(x)$$
$$= \int_\mathbb{R} v(x - t) \mathrm{d}P \circ X_2^{-1}(x) = \int_\Omega v(X_2(\omega) - t) \mathrm{d}P(\omega) = \mathbb{E}[v(X_2 - t)].$$

Hence, $\mathrm{AVaR}_\beta^\varepsilon$ and σ_β^ε are law invariant. As a result, a large class of risk measures/optimized certainty equivalents are law invariant.

References

[1] Aliprantis CD, Border KC (2006) *Infinite Dimensional Analysis: A Hitchhiker's Guide* (Berlin: Springer), 3rd edition, ISBN 978-3-540-32696-0; 3-540-32696-0, URL http://dx.doi.org/10.1007/3-540-29587-9.

[2] Artzner P, Delbaen F, Eber JM, Heath D (1999) Coherent measures of risk. *Math. Finance* 9(3):203–228, URL http://dx.doi.org/10.1111/1467-9965.00068.

[3] Attouch H (1984) *Variational Convergence for Functions and Operators.* Applicable Mathematics Series (Boston, MA: Pitman), ISBN 0-273-08583-2.

[4] Attouch H, Wets RJB (1989) Epigraphical analysis. *Ann. Inst. H. Poincaré Anal. Non Linéaire* 6(suppl.):73–100, ISSN 0294-1449, URL http://dx.doi.org/10.1016/S0294-1449(17)30036-7.

[5] Aubin JP, Frankowska H (2009) *Set-Valued Analysis.* Mod. Birkhäuser Class. (Boston, MA: Springer), URL http://dx.doi.org/10.1007/978-0-8176-4848-0.

[6] Beiser F, Keith B, Urbainczyk S, Wohlmuth B (2020) Adaptive sampling strategies for risk-averse stochastic optimization with constraints. *arXiv preprint arXiv:2012.03844* URL http://arxiv.org/abs/2012.03844.

[7] Ben-Tal A, Teboulle M (2007) An old-new concept of convex risk measures: The optimized certainty equivalent. *Math. Finance* 17(3):449–476, URL http://dx.doi.org/10.1111/j.1467-9965.2007.00311.x.

[8] Bendsøe MP, Sigmund O (2003) *Topology Optimization: Theory, Methods and Applicatons* (Berlin: Springer), URL http://dx.doi.org/10.1007/978-3-662-05086-6.

[9] Berner J, Grohs P, Jentzen A (2020) Analysis of the generalization error: empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *SIAM J. Math. Data Sci.* 2(3):631–657, URL http://dx.doi.org/10.1137/19M125649X.

[10] Billingsley P (2012) *Probability and Measure.* Wiley Ser. Probab. Stat. (Hoboken, NJ: John Wiley & Sons).

[11] Bogachev VI (2007) *Measure Theory* (Berlin: Springer), URL http://dx.doi.org/10.1007/978-3-540-34514-5.

[12] Bonnans JF, Shapiro A (2000) *Perturbation Analysis of Optimization Problems*. Springer Ser. Oper. Res. (New York: Springer), URL http://dx.doi.org/10.1007/978-1-4612-1394-9.

[13] Borwein JM, Lewis AS (1994) Strong rotundity and optimization. *SIAM J. Optim.* 4(1):146–158, ISSN 1052-6234, URL http://dx.doi.org/10.1137/0804008.

[14] Borwein JM, Vanderwerff JD (2010) *Convex functions: constructions, characterizations and counterexamples*. Encyclopedia Math. Appl. 109 (Cambridge University Press, Cambridge), ISBN 978-0-521-85005-6, URL http://dx.doi.org/10.1017/CBO9781139087322.

[15] Curi S, Levy KY, Jegelka S, Krause A (2020) Adaptive sampling for stochastic risk-averse learning. Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H, eds., *Advances in Neural Information Processing Systems*, volume 33, 1036–1047 (Curran Associates, Inc.), URL https://proceedings.neurips.cc/paper/2020/file/0b6ace9e8971cf36f1782aa982a708db-Paper.pdf.

[16] Davis D, Drusvyatskiy D (2022) Graphical convergence of subgradients in nonconvex optimization and learning. *Math. Oper. Res.* 47(1):209–231, ISSN 0364-765X, URL http://dx.doi.org/10.1287/moor.2021.1126.

[17] Dong MX, Wets RJB (2000) Estimating density functions: a constrained maximum likelihood approach. *J. Nonparametr. Statist.* 12(4):549–595, URL http://dx.doi.org/10.1080/10485250008832822.

[18] Dudley RM (2002) *Real Analysis and Probability*, volume 74 of *Cambridge Studies in Advanced Mathematics* (Cambridge University Press, Cambridge), ISBN 0-521-00754-2, URL http://dx.doi.org/10.1017/CBO9780511755347.

[19] Dudley RM, Norvaiša R (2011) *Concrete Functional Calculus*. Springer Monographs in Mathematics (New York: Springer), ISBN 978-1-4419-6949-1, URL http://dx.doi.org/10.1007/978-1-4419-6950-7.

[20] Dupačová J, Wets RJB (1988) Asymptotic behavior of statistical estimators and of optimal solutions of stochastic optimization problems. *Ann. Statist.* 16(4):1517–1549, URL http://dx.doi.org/10.1214/aos/1176351052.

[21] Föllmer H, Schied A (2002) Convex measures of risk and trading constraints. *Finance Stoch.* 6(4):429–447, ISSN 0949-2984, URL http://dx.doi.org/10.1007/s007800200072.

[22] Garreis S, Surowiec TM, Ulbrich M (2021) An interior-point approach for solving risk-averse PDE-constrained optimization problems with coherent risk measures. *SIAM J. Optim.* 31(1):1–19, URL http://dx.doi.org/10.1137/19M125039X.

[23] Garreis S, Ulbrich M (2019) A fully adaptive method for the optimal control of semilinear elliptic PDEs under uncertainty using low-rank tensors. Preprint, Technische Universität München, München, URL http://go.tum.de/204409.

[24] Geiersbach C, Scarinci T (2021) Stochastic proximal gradient methods for nonconvex problems in Hilbert spaces. *Comput. Optim. Appl.* 78(3):705–740, URL http://dx.doi.org/10.1007/s10589-020-00259-y.

[25] Gwinner J, Jadamba B, Khan AA, Raciti F (2022) *Uncertainty Quantification in Variational Inequalities: Theory, Numerics, and Applications* (Boca Raton, FL: CRC), URL http://dx.doi.org/10.1201/9781315228969.

[26] Hertlein L, Rauls AT, Ulbrich M, Ulbrich S (2022) An inexact bundle method and subgradient computations for optimal control of deterministic and stochastic obstacle problems. Hintermüller M, Herzog R, Kanzow C, Ulbrich M, Ulbrich S, eds., *Non-Smooth and Complementarity-Based Distributed Parameter Systems: Simulation and Hierarchical Optimization*, 467–497, Internat. Ser. Numer. Math. 172 (Cham: Birkhäuser), URL http://dx.doi.org/10.1007/978-3-030-79393-7_19.

[27] Hille E, Phillips RS (1974) *Functional Analysis and Semi-Groups*. Colloq. Publ. 31 (Providence, RI: AMS), URL http://dx.doi.org/10.1090/coll/031.

[28] Hoffhues M, Römisch W, Surowiec TM (2021) On quantitative stability in infinite-dimensional optimization under uncertainty. *Optim. Lett.* 15(8):2733–2756, ISSN 1862-4472, URL http://dx.doi.org/10.1007/s11590-021-01707-2.

[29] Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. Le Cam LM, Neyman J, eds., *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 221–233 (Berkeley, CA: University of California Press).

[30] Juditsky A, Kwon J, Moulines E (2022) Unifying mirror descent and dual averaging. *Math. Program.* URL http://dx.doi.org/10.1007/s10107-022-01850-3.

[31] King AJ, Rockafellar RT (1993) Asymptotic theory for solutions in statistical estimation and stochastic programming. *Math. Oper. Res.* 18(1):148–162, ISSN 0364-765X, URL http://dx.doi.org/10.1287/moor.18.1.148.

[32] Kleywegt AJ, Shapiro A, Homem-de Mello T (2002) The sample average approximation method for stochastic discrete optimization. *SIAM J. Optim.* 12(2):479–502, URL http://dx.doi.org/10.1137/S1052623499363220.

[33] Korf LA, Wets RJB (2001) Random LSC Functions: An Ergodic Theorem. *Math. Oper. Res.* 26(2):421–445, URL http://dx.doi.org/10.1287/moor.26.2.421.10548.

[34] Kouri DP, Surowiec TM (2016) Risk-averse PDE-constrained optimization using the conditional value-at-risk. *SIAM J. Optim.* 26(1):365–396, URL http://dx.doi.org/10.1137/140954556.

[35] Kouri DP, Surowiec TM (2018) Existence and optimality conditions for risk-averse PDE-constrained optimization. *SIAM/ASA J. Uncertain. Quantif.* 6(2):787–815, URL http://dx.doi.org/10.1137/16M1086613.

[36] Kouri DP, Surowiec TM (2020) Epi-regularization of risk measures. *Math. Oper. Res.* 45(2):774–795, URL http://dx.doi.org/10.1287/moor.2019.1013.

[37] Kouri DP, Surowiec TM (2020) Risk-averse optimal control of semilinear elliptic PDEs. *ESAIM Control. Optim. Calc. Var.* 26(53), URL http://dx.doi.org/10.1051/cocv/2019061.

[38] Lachout P, Liebscher E, Vogel S (2005) Strong convergence of estimators as ϵ_n-minimisers of optimisation problems. *Ann. Inst. Statist. Math.* 57(2):291–313, ISSN 0020-3157, URL http://dx.doi.org/10.1007/BF02507027.

[39] Lazarov BS, Sigmund O (2011) Filters in topology optimization based on Helmholtz-type differential equations. *Internat. J. Numer. Methods Engrg.* 86(6):765–781, ISSN 0029-5981, URL http://dx.doi.org/10.1002/nme.3072.

[40] Liu J, Cui Y, Pang JS (2022) Solving nonsmooth and nonconvex compound stochastic programs with applications to risk measure minimization. *Math. Oper. Res.* URL http://dx.doi.org/10.1287/moor.2021.1247.

[41] Mei S, Bai Y, Montanari A (2018) The landscape of empirical risk for nonconvex losses. *Ann. Statist.* 46(6A):2747–2774, ISSN 0090-5364, URL http://dx.doi.org/10.1214/17-AOS1637.

[42] Milz J (2022) Consistency of Monte Carlo estimators for risk-neutral PDE-constrained optimization. *arXiv preprint arXiv:2204.04809* URL http://arxiv.org/abs/2204.04809.

[43] Milz J (2022) Reliable error estimates for optimal control of linear elliptic PDEs with random inputs. *arXiv preprint arXiv:2206.09160* URL https://arxiv.org/abs/2206.09160.

[44] Milz J (2022) Sample average approximations of strongly convex stochastic programs in Hilbert spaces. *Optim. Lett.* URL http://dx.doi.org/10.1007/s11590-022-01888-4.

[45] Mosco U (1969) Convergence of convex sets and of solutions of variational inequalities. *Advances in Mathematics* 3(4):510–585, ISSN 0001-8708, URL http://dx.doi.org/10.1016/0001-8708(69)90009-7.

[46] Nelsen NH, Stuart AM (2021) The random feature model for input-output maps between Banach spaces. *SIAM J. Sci. Comput.* 43(5):A3212–A3243, ISSN 1064-8275, URL http://dx.doi.org/10.1137/20M133957X.

[47] Pflug GCh (1995) Asymptotic stochastic programs. *Math. Oper. Res.* 20(4):769–789, URL http://dx.doi.org/10.1287/moor.20.4.769.

[48] Pflug GCh (2003) Stochastic optimization and statistical inference. *Stochastic Programming*, 427 – 482, Handbooks Oper. Res. Manag. Sci. 10 (Elsevier), URL http://dx.doi.org/10.1016/S0927-0507(03)10007-2.

[49] Qi Z, Cui Y, Liu Y, Pang JS (2021) Asymptotic properties of stationary solutions of coupled nonconvex nonsmooth empirical risk minimization. *Math. Oper. Res.* URL http://dx.doi.org/10.1287/moor.2021.1198.

[50] Rachev ST, Römisch W (2002) Quantitative stability in stochastic programming: the method of probability metrics. *Math. Oper. Res.* 27(4):792–818, ISSN 0364-765X, URL http://dx.doi.org/10.1287/moor.27.4.792.304.

[51] Robinson SM (1996) Analysis of sample-path optimization. *Math. Oper. Res.* 21(3):513–528, ISSN 0364-765X, URL http://dx.doi.org/10.1287/moor.21.3.513.

[52] Römisch W, Schultz R (1991) Distribution sensitivity in stochastic programming. *Math. Program.* 50(2, (Ser. A)):197–226, ISSN 0025-5610, URL http://dx.doi.org/10.1007/BF01594935.

[53] Römisch W, Surowiec TM (2021) Asymptotic properties of Monte Carlo methods in elliptic PDE-constrained optimization under uncertainty. *arXiv preprint arXiv:2106.06347* URL https://arxiv.org/abs/2106.06347.

[54] Royset JO (2020) Approximations of semicontinuous functions with applications to stochastic optimization and statistical estimation. *Math. Program.* 184:289–318, ISSN 0025-5610, URL http://dx.doi.org/10.1007/s10107-019-01413-z.

[55] Royset JO, Wets RJB (2020) Variational analysis of constrained M-estimators. *Ann. Statist.* 48(5):2759–2790, ISSN 0090-5364, URL http://dx.doi.org/10.1214/19-AOS1905.

[56] Ruszczyński A, Shapiro A (2006) Optimization of Convex Risk Functions. *Math. Oper. Res.* 31(3):433–452, URL http://dx.doi.org/10.1287/moor.1050.0186.

[57] Shapiro A (1989) Asymptotic properties of statistical estimators in stochastic programming. *Ann. Statist.* 17(2):841–858, ISSN 0090-5364, URL http://dx.doi.org/10.1214/aos/1176347146.

[58] Shapiro A (1991) Asymptotic analysis of stochastic programs. *Ann. Oper. Res.* 30(1):169–186, URL http://dx.doi.org/10.1007/BF02204815.

[59] Shapiro A (1993) Asymptotic behavior of optimal solutions in stochastic programming. *Math. Oper. Res.* 18(4):829–845, URL http://dx.doi.org/10.1287/moor.18.4.829.

[60] Shapiro A (2000) Statistical Inference of Stochastic Optimization Problems. Uryasev SP, ed., *Probabilistic Constrained Optimization: Methodology and Applications*, 282–307, Probabilistic Constrained Optimization (Boston, MA: Springer), URL http://dx.doi.org/10.1007/978-1-4757-3150-7_16.

[61] Shapiro A (2013) Consistency of sample estimates of risk averse stochastic programs. *J. Appl. Probab.* 50(2):533–541, URL http://dx.doi.org/10.1239/jap/1371648959.

[62] Shapiro A (2013) On Kusuoka representation of law invariant risk measures. *Math. Oper. Res.* 38(1):142–152, ISSN 0364-765X, URL http://dx.doi.org/10.1287/moor.1120.0563.

[63] Shapiro A, Dentcheva D, Ruszczyński A (2021) *Lectures on Stochastic Programming: Modeling and Theory*. MOS-SIAM Ser. Optim. (Philadelphia, PA: SIAM), 3rd edition, URL http://dx.doi.org/10.1137/1.9781611976595.

[64] Shapiro A, Nemirovski A (2005) On complexity of stochastic programming problems. Jeyakumar V, Rubinov A, eds., *Continuous Optimization: Current Trends and Modern Applications*, 111–146, Appl. Optim. 99 (Boston, MA: Springer), URL http://dx.doi.org/10.1007/0-387-26771-9_4.

[65] Sigmund O, Maute K (2013) Topology optimization approaches. *Structural and Multidisciplinary Optimization* 48(6):1031–1055, ISSN 1615-1488, URL http://dx.doi.org/10.1007/s00158-013-0978-6.

[66] Surowiec TM (2018) Numerical optimization methods for the optimal control of elliptic variational inequalities. Antil H, Kouri DP, Lacasse MD, Ridzal D, eds., *Frontiers in PDE-Constrained Optimization*, 123–170, IMA Vol. Math. Appl. 163 (New York, NY: Springer), ISBN 978-1-4939-8636-1, URL http://dx.doi.org/10.1007/978-1-4939-8636-1_4.

[67] Wechsung F, Giuliani A, Landreman M, Cerfon AJ, Stadler G (2022) Single-stage gradient-based stellarator coil design: stochastic optimization. *Nuclear Fusion* 62(7):076034, URL http://dx.doi.org/10.1088/1741-4326/ac45f3.

[68] Zeidler E (1990) *Nonlinear Functional Analysis and its Applications* II/B: Nonlinear Monotone Operators (New York: Springer), ISBN 0-387-97167-X, URL http://dx.doi.org/10.1007/978-1-4612-0981-2.

[69] Zervos M (1999) On the epiconvergence of stochastic optimization problems. *Math. Oper. Res.* 24(2):495–508, ISSN 0364-765X, URL http://dx.doi.org/10.1287/moor.24.2.495.