Abstract. Stochastic ambiguity provides a rich class of uncertainty models that includes those in stochastic, robust, risk-based, and semi-infinite optimization and that accounts for uncertainty about parameter values as well as incompleteness of the description of uncertainty. We provide a novel, unifying perspective on optimization under stochastic ambiguity that rests on two pillars. First, ambiguity is formulated in terms of the (cumulative) probability distribution associated with the random elements; more specifically, ambiguity is expressed by letting this distribution belong to a subfamily of distributions that might, or might not, depend on the decision variable. We derive a series of estimates by introducing a metric for the space of distribution functions based on the hypo-distance between upper semicontinuous functions. In the process, we show that this metric is consistent with convergence in distribution (weak convergence) of the associated probability measures. Second, we rely on the theory of lopsided convergence to establish existence, convergence, and approximation of solutions of optimization problems with stochastic ambiguity. For the first time, we estimate a distance between bifunctions and show that this leads to bounds on the solution quality for problems with stochastic ambiguity. Among other consequences, these results facilitate the study of the “price of robustness” and related quantities.

Key words. stochastic ambiguity, robust optimization, lopsided convergence, lop-distance, rate of convergence, price of robustness, weak convergence

AMS subject classifications. 90C15, 90C34, 90C47, 65K10

DOI. 10.1137/16M1060704

1. Introduction. Optimization under stochastic ambiguity accounts for uncertainty in parameters as well as the fact that models of uncertainty might also be imprecise. The goal is to choose the decision variables so as to minimize the worst-case value of a cost function given a family of potential probability measures. A probability measure models the uncertainty about parameter values, and the set captures the ambiguity about the correct measure. Optimization under stochastic ambiguity includes as special cases robust [4, 5, 15], stochastic [7, 21, 50], semi-infinite [27, 17, 51], and risk-based [25, 47, 34, 50] optimization. The references provide a glimpse into a vast literature, where applications in finance are prevalent; see, for example, [9, 24].

In this paper, stochastic ambiguity is formulated in terms of (cumulative) distribution functions, bypassing the need to impose the usual restrictions about integrable functions. We exploit and refine the variational theory for upper semicontinuous functions, which eventually allows us to provide a novel approach to approximation and convergence results for distribution functions while providing an alternative perspective on some classical results from probability theory. From there, we introduce the variational theory of lopsided convergence to the problem class and show how it enables us to derive the existence, convergence, and approximation of solutions. The
development applies to general metric spaces and, thus, to nearly every formulation of stochastic ambiguity problems, including those using distribution functions in the manner laid out here and those with decision-dependent ambiguity sets. Detailed results about errors in solutions require us to develop for the first time estimates of the lop-distance, which quantifies lopsided convergence.

The formulation of ambiguity sets in terms of distribution functions on $\mathbb{R}^m$, instead of probability measures, leads to a treatment of such functions as a subset of the metric space of upper semicontinuous functions equipped with the hypo-distance. For the first time, we define convexity of subsets of such functions and provide estimates of the hypo-distance for distribution functions. Moreover, we provide an explicit statement of the fact that convergence of distribution functions in the hypo-distance is equivalent to weak convergence, with a new proof relying exclusively on elementary concepts from variational analysis; the result is implicit in [46, 45]. Of course, there are numerous metrics available for spaces of probability measures including the Levy–Prokhorov and bounded Lipschitz metrics, which also characterize weak convergence in the present context; the Wasserstein metrics, which require finite moments and are stronger than weak convergence; and the total variation metric, which is also stronger than weak convergence. This paper is the first attempt to use the hypo-distance as a metric for distribution functions in the context of optimization under stochastic ambiguity. As we see below, it is a promising alternative because of its natural interpretation, its metrization of the topology of weak convergence, the compact balls in this metric (a property difficult to ascertain in infinite dimensions), and the fact that it requires no moment assumptions. Moreover, the formulation can easily handle any finite dimension, i.e., any number of uncertain parameters. In fact, its extension to infinite-dimensional spaces appears clear (see the ideas in [45]), but such possibilities are beyond the scope of the present article.

The subject of stability of stochastic programs under distributional perturbations relates to problems with stochastic ambiguity, as there also perturbations in probability measures need to be considered. Naturally, as stochastic programs involve integral functions, the vast majority of such stability results focus on stronger metrics than those metrizing the topology of weak convergence; see, for example, [13, 36, 37], which provide an extensive treatment of stochastic programming stability, including discussion of a large number of metrics on spaces of probability measures and general Lipschitz continuity results in the sense of $\zeta$-structures for probability measures. Theorem 6 of [36] also briefly considers approximations in the sense of weak convergence in the context of stochastic programs, but results there do not reach quantitative solution estimates. For two-stage stochastic programs with risk measures, [26] develops quantitative results under $\zeta$-structures. More generally, stability of optimization problems is treated comprehensively in [22, 8, 35].

Using the total variation metric, a class of problems with stochastic ambiguity is examined in [52] and a recent extension [53]. The latter paper is closely related to our study in that it also considers decision-dependent ambiguity and, in fact, motivates such dependence very well. However, there are three main differences between that work and our paper. (i) [53] deals with ambiguity sets given in terms of integral functions such as moments. A large portion of our paper takes a more general approach and in many examples considers ambiguity sets given in terms of constraints acting directly on distribution functions. (ii) [53] adopts the total variation metric, which is stronger than our choice of the hypo-distance. (iii) [53] provides local results, i.e., considers small perturbations in the probability measure away from that of an original problem. We give global results and essentially permit any size of “perturbation.”
The variational theory of lopsided convergence of bifunctions (a convenient abbreviation for bivariate functions, i.e., those depending on a pair of variables) is ideally suited for examining the convergence of approximations of optimization problems with stochastic ambiguity. The notion originated with [2] and later was modified and extended in [18, 19, 43, 41]. Its application is not limited to any one formulation of stochastic ambiguity, and it applies to all metric spaces. Although we focus on formulations in terms of distribution functions, we demonstrate other possibilities as well.

For the first time, we estimate the lop-distance between bifunctions, a quantification of lopsided convergence proposed in [41]. Utilizing these estimates, the paper bounds the difference between optimal solutions and optimal values of two optimization problems with stochastic ambiguity. The results provide a new quantitative theory of approximation for such problems.

Throughout, our aim is to keep assumptions to a minimum. In particular, we do not universally insist that the feasible sets of decision variables and the ambiguity sets be compact, though sometimes the latter turn out to be compact under natural assumptions when formulations are in terms of distribution functions. Neither do we require convexity and/or concavity. We refer to the extensive literature on robust, stochastic, and risk-based optimization for specialized results; see references above and the next section.

The paper proceeds as follows: in section 2 we give problem statements, illustrative examples, and background material. Section 3 develops the foundations for studying distribution functions in the present context. Section 4 reviews the notion of lopsided convergence, states the main consequences, and provides specific results for optimization under stochastic ambiguity. Section 5 offers estimates of the lop-distance and solution quality.

2. Problem formulation and examples. Although extensions are possible, and in fact trivial in some cases, we limit the scope of our analysis to optimization of an n-dimensional vector of decision variables in the presence of uncertainty about m parameters. This enables us to skirt much of the topological and other technical considerations needed in the more general cases. The set of all probability measures on \((\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})\) is denoted by \(\mathcal{M}\), where \(\mathcal{B}_{\mathbb{R}^m}\) is the Borel \(\sigma\)-algebra on \(\mathbb{R}^m\). Given a real-valued cost function \(\varphi\) defined on a subset of \(\mathbb{R}^n \times \mathcal{M}\), the problem of optimization under stochastic ambiguity takes the form

\[
\min_{x \in C \subset \mathbb{R}^n} \sup_{P \in \mathcal{M}_0(x)} \varphi(x, P),
\]

where the ambiguity set \(\mathcal{M}_0(x) \subset \mathcal{M}\) might depend on the \(x\)-decision.

2.1. Ambiguity problem and examples. The problem of optimization under stochastic ambiguity provides practitioners the ability to model decision making under uncertainty while accounting for possibly incomplete, uncertain, and decision-dependent descriptions of unknown parameters. Three examples illustrate some possibilities.

Example 1: Expectation minimization. Suppose that \(\varphi(x, P) = \int \psi(x, \xi) dP(\xi)\) for some \(\psi : C \times \mathbb{R}^m \to \mathbb{R}\) that, for all \(x \in C \subset \mathbb{R}^n\), is \(P\)-integrable in its second argument for every \(P\) in some nonempty subset \(\mathcal{M}_0(x) \subset \mathcal{M}\). This case captures stochastic optimization under distributional uncertainty (most often considered with an ambiguity set \(\mathcal{M}_0(x)\) that is independent of \(x\)), stochastic programs with decision-
dependent probability measure where $\mathcal{M}_0(x)$ is a singleton for every $x$, and certain Stackelberg games; see, for example, [20, 23, 31, 48]. In particular, we refer the reader to [14] for an overview of ambiguity models, especially those from generalized moment constraints and unimodality restrictions, and to [10] for an extensive discussion of partial knowledge about first- and second-order moments and the associated computability. A particular class of ambiguity sets $\mathcal{M}_0$ (independent of $x$) that permits a reformulation as a stochastic program is developed in [49].

Example 2: Risk minimization. For some $m$-dimensional random vector $\xi$ with distribution $P_0$, suppose that the function $\psi : C \times \mathbb{R}^m \to \mathbb{R}$ defines at each $x \in C$ a random variable $\psi(x, \xi)$. It is well known that a coherent measure of risk $\mathcal{R}$ defined on a subset of random variables [1, 34] can be written in the form $\mathcal{R}(\psi(x, \xi)) = \sup_{P \in \mathcal{M}_0} \int \psi(x, \xi) dP(\xi)$, provided that $\psi(x, \xi)$ belongs to the domain of $\mathcal{R}$. Here, $\mathcal{M}_0$ is the set of probability measures on $\mathbb{R}^m$ absolutely continuous with respect to $P_0$. Each measure has a density in the domain of the conjugate of $\mathcal{R}$; see, for example, [44, 16, 34, 31]. Consequently, the problem is cast in the same form as Example 1, but with an ambiguity set that is independent of $x$. If the decision $x$ influences the choice of risk measure, an interesting possibility which has not received much attention, then $\mathcal{M}_0$ remains dependent on $x$. Superquantile risk measures\(^1\) featured in [33, 44, 32] are particular instances of such risk measures most naturally defined on the space $\mathcal{L}^1(\mathbb{R}^m) := \{g : \mathbb{R}^m \to \mathbb{R} : \int |g(\xi)| dP(\xi) < \infty\}$ of integrable random variables. Specifically, suppose that the random variable $\psi(x, \xi) \in \mathcal{L}^1(\mathbb{R}^m)$ and $\alpha \in [0, 1)$; then the $\alpha$-superquantile risk measure is

$$\mathcal{R}_{\alpha}(\psi(x, \xi)) := \sup_{q \in \mathcal{L}^\infty(\mathbb{R}^m)} \left\{ \int \psi(x, \xi) q(\xi) dP(\xi) : 0 \leq q(\xi) \leq \frac{1}{1-\alpha} \text{ for } P_0\text{-a.e. } \xi \in \mathbb{R}^m, \int q(\xi) dP(\xi) = 1 \right\},$$

where $\mathcal{L}^\infty(\mathbb{R}^m) := \{g : \mathbb{R}^m \to \mathbb{R} : g \text{ measurable, essentially bounded}\}$. Thus, superquantile minimization amounts to optimization under stochastic ambiguity with the inner max-problem being subsets of $\mathcal{L}^\infty(\mathbb{R}^m)$.

Example 3: Robust optimization. For some set-valued mapping $\Xi : C \to \mathbb{R}^m$, with $\Xi(x) \neq \emptyset$ for all $x \in C$, let $\mathcal{M}_0(x) = \{P \in \mathcal{M} : P(\xi) = 1 \text{ for some } \xi \in \Xi(x)\}$. Then, optimization under stochastic ambiguity in the form of Example 1 leads to the problem $\min_{x \in C} \sup_{\Xi \in \Xi(x)} \psi(x, \xi)$, which encapsulates many robust optimization and (generalized) semi-infinite programming problems; see, for example, [27, 51, 4, 5, 15].

It is apparent that optimization problems with stochastic ambiguity come in a wide variety of forms. Although they might be stated with an ambiguity set of probability measures, as in Example 1, convenient formulations in terms of ambiguity sets of random variables (Example 2) and parameter vectors (Example 3) arise as well. In the next subsection, we give a novel formulation in terms of distribution functions viewed as elements of a space of upper semicontinuous functions. To permit a unified treatment of the vast number of possibilities, throughout this work we consider ambiguity formulated as a subset of a general metric space $(Y, d_Y)$, which captures all of the above possibilities and many more, and thus define the ambiguity problem

(AP): $\min_{x \in C} \sup_{y \in D(x)} \phi(x, y)$,
where the real-valued bifunction

$$\varphi \in \text{bfcns}(\mathbb{R}^n, Y) := \{ \varphi : \text{dom} \varphi \to \mathbb{R} : \emptyset \neq \text{dom} \varphi \subset \mathbb{R}^n \times Y \}$$

has a nonempty domain characterized by a set

$$C := \{ x \in \mathbb{R}^n : \exists y \in Y \text{ such that } (x, y) \in \text{dom} \varphi \}$$

and the set-valued mapping $D : C \rightrightarrows Y$ such that

$$D(x) := \{ y \in Y : (x, y) \in \text{dom} \varphi \} \text{ for } x \in C.$$ 

Consequently, $\text{dom} \varphi = \{ (x, y) \in \mathbb{R}^n \times Y : x \in C, \ y \in D(x) \}$, with $\text{dom} D = \{ x \in \mathbb{R}^n : D(x) \neq \emptyset \} = C$. Given $\varphi \in \text{bfcns}(\mathbb{R}^n, Y)$, it is understood that $C$ and $D$ are the corresponding set and set-valued mapping, respectively, that specify the domain of the bifunction. With a slight abuse of notation, occasionally we simply write $D$ instead of $D(x)$ when the set-valued mapping is constant on its domain. Under decision $x$, $D(x)$ is the ambiguity set.

Applications might specify a bifunction to be optimized, possibly defined on a large domain, and separately specify constraints and ambiguity sets. However, it is trivial to restrict such a bifunction to a domain given by the constraints and ambiguity sets and thereby return to the present context where a bifunction $\varphi \in \text{bfcns}(\mathbb{R}^n, Y)$ fully determines an instance of (AP).

An $x$ that achieves the minimum in (AP) is called a minsup-point of (AP), and the collection of all such points is denoted by $\text{argminsup} \varphi$. The optimal value of (AP) is denoted by $\text{minsup} \varphi$ and is called its minsup-value. The minsup-points and minsup-value of (AP) are also called the minsup-points and minsup-value, respectively, of $\varphi$.

### 2.2. Formulation using distribution functions.

We here explore for the first time a formulation of stochastic ambiguity centered on cumulative distribution functions viewed as elements of a space of upper semicontinuous functions on $\mathbb{R}^m$. Practitioners usually think about uncertainty in terms of a random vector $\xi$ with some distribution function that is more or less known. Consequently, it is natural to formulate optimization under stochastic ambiguity in terms of distribution functions that must be selected from a set of “plausible” distributions. A geometrically intuitive metric, the hypo-distance to be defined below, emerges to quantify the distance between two distribution functions. The metric has the convenient property that closed balls in the space are compact, which facilitates several results including those related to existence of solutions. Since the convergence induced by the hypo-distance metric is equivalent to weak convergence of distribution functions, we do not have to rely on the existence of moments.

We start by recalling some well-known facts about distribution functions on $\mathbb{R}^m$. Every probability measure $P$ on $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ defines a distribution function $F : \mathbb{R}^m \to [0, 1]$ through $F(\xi) = P(S_{\xi})$ for $\xi \in \mathbb{R}^m$, where $S_{\xi} := \{ \zeta \in \mathbb{R}^m : \zeta \leq \xi \}$. Vector inequalities are understood componentwise. A distribution function $F$ is non-decreasing, i.e., $F(\zeta) \leq F(\xi)$ for $\zeta \leq \xi$; it is continuous from above;\(^2\) it satisfies $F(\xi^i) \to 0$ whenever one of the components of $\xi^i$ tends to $-\infty$, with the others held fixed, and $F(\xi^i) \to 1$ if $\xi^i \to \infty$ for all $i$; and it has $\Delta_A F \geq 0$ for every rectangle$^3$ $A$, $P(\xi^i = a_i \lor \zeta^i \downarrow 0 \text{ for all } i = 1, 2, \ldots, m.$ $^3$A rectangle in $\mathbb{R}^m$ is of the form $A = \{ \xi : a_i < \xi_i \leq b_i, i = 1, 2, \ldots, m \}$ for real $a_i$ and $b_i$.\(^2\)
where $\Delta_A F := \sum_{j=1}^{2^m} (\text{sgn}_A v^j) F(v^j)$, with $v^j$, $j = 1, \ldots, 2^m$, being the vertices of $A$, and $\text{sgn}_A v^j = 1$ if the number of components $v^j_i$ at a lower bound of $A$ is even and $\text{sgn}_A v^j = -1$ if the number is odd. (For $m = 2$, $\Delta_A F = F(v^1) - F(v^2) - F(v^3) + F(v^4)$, where $v^1$ is the upper right vertex, $v^2$ and $v^3$ are the upper left and lower right vertices, and $v^4$ is the lower left vertex.)

For every $F : \mathbb{R}^m \to \mathbb{R}$ with these properties (nondecreasing, continuity from above, limits, $\Delta_A F \geq 0$), there exists a unique probability measure $P$ on $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ such that $P(A) = \Delta_A F$ for rectangles $A$, and $P(S_\xi) = F(\xi)$ for all $\xi \in \mathbb{R}^m$. We denote by

$$\text{cd-fcns}(\mathbb{R}^m) := \{ F : \mathbb{R}^m \to [0, 1] : \exists P \in \mathcal{M} \text{ with } F(\xi) = P(S_\xi) \ \forall \xi \in \mathbb{R}^m \}$$

the set of (cumulative) distribution functions. Formulation of (AP) in terms of distribution functions then amounts to specifying ambiguity sets that are subsets of $\text{cd-fcns}(\mathbb{R}^m)$.

To facilitate analysis of distribution functions, we view them as subsets of the larger class $\text{usc-fcns}(\mathbb{R}^m; [0, 1])$ of upper semicontinuous (usc) functions $g : \mathbb{R}^m \to [0, 1]$; i.e., for every $\xi^* \to \xi \in \mathbb{R}^m$, $\limsup g(\xi^*) \leq g(\xi)$. If $F \in \text{cd-fcns}(\mathbb{R}^m)$, $\xi^* \to \xi$, and $\xi^*_i = \max\{\xi_i, \xi^*_i\}$ for all $i$, then $F(\xi^*_i) \leq F(\xi^*_i)$ by the nondecreasing property of $F$. Since $F(\xi^*_i) \to F(\xi)$ by the continuity-from-above property, it is clear that $F$ is usc, and thus, $\text{cd-fcns}(\mathbb{R}^m) \subset \text{usc-fcns}(\mathbb{R}^m; [0, 1])$.

We embed the space $\text{usc-fcns}(\mathbb{R}^m; [0, 1])$ with the hypothesis $d^h$, which quantifies the distance between usc functions in terms of a distance between their hypographs. The hypograph of $g \in \text{usc-fcns}(\mathbb{R}^m; [0, 1])$ is hypo $g := \{(\xi, \xi_0) \in \mathbb{R}^m \times \mathbb{R} : g(\xi) \geq \xi_0 \}$. For $\mathbb{R}^m \times \mathbb{R}$ we adopt the norm

$$\|(\xi, \xi_0)\| := \max\{\|\xi\|_2, |\xi_0|\} \text{ for } (\xi, \xi_0) \in \mathbb{R}^m \times \mathbb{R}. \tag{1}$$

This choice and the restriction to usc functions with values in $[0, 1]$ represent minor deviations from the approach taken in [35, Chapters 4 and 7], but all the major properties remain the same. The choice of norm is motivated by a simplification in bounds on the hypo-distance and future extensions to situations when $\mathbb{R}^m$ is replaced by a general metric space. A ball $S(\xi, r) := \{\tilde{\xi} \in \mathbb{R}^m \times \mathbb{R} : \|\tilde{\xi} - \xi\| \leq r \}$ for $\xi = (\xi, \xi_0) \in \mathbb{R}^m \times \mathbb{R}$ is actually a "hyper-cylinder," as hinted at by the symbol $S$. Figure 1 shows $S(0, \rho)$ for the case $m = 1$. The distance between $\xi \in \mathbb{R}^m \times \mathbb{R}$ and a set $S \subset \mathbb{R}^m \times \mathbb{R}$ is given by dist $(\xi, S) := \inf\{\|\xi - \tilde{\xi}\|_S : \tilde{\xi} \in S\}$.

We are then in a position to define, for any $g, g' \in \text{usc-fcns}(\mathbb{R}^m; [0, 1])$, the hypo-
distance as
\[ d^h(g, g') := \int_0^\infty d^h_\rho(g, g')e^{-\rho}d\rho, \]
where the \( \rho \)-hypodistance is
\[ d^h_\rho(g, g') := \max \left\{ \left| \text{dist} \left( \xi, \text{hypo} \ g \right) - \text{dist} \left( \xi, \text{hypo} \ g' \right) \right| : \|\xi\|_S \leq \rho \} \text{ for } \rho \geq 0. \]

Figure 1 illustrates the situation and especially the hypo-distance between distribution functions \( F \) and \( G \). We observe that \( d^h_\rho \) is not identical to the classical Pompeiu–Hausdorff distance, which anyways applies only to compact sets; see [35, Exercise 7.60, Proposition 7.61]. In the next section, we give specific estimates of the hypodistance between distribution functions beyond the immediate fact that \( d^h(g, g') \leq 1 \) for all \( g, g' \in \text{usc -fcns}(\mathbb{R}^m; [0, 1]) \).

The hypodistance is a metric on the space \( \text{usc -fcns}(\mathbb{R}^m; [0, 1]) \) and induces the \textit{hypo-topology} (sometimes called the Attouch–Wets topology). In fact, we deduce from [35, Theorem 7.58] and [42, Corollary 3.6] that \( \{ \text{usc -fcns}(\mathbb{R}^m; [0, 1]), d^h \} \) is a complete separable metric (Polish) space. Every ball \( \{ g' \in \text{usc -fcns}(\mathbb{R}^m; [0, 1]) : d^h(g, g') \leq r \} \) in this space is compact, as can be deduced from [35, Theorem 7.58]. We say that \( g' \in \text{usc -fcns}(\mathbb{R}^m; [0, 1]) \) \textit{hypoco}nverges to a function \( g \in \text{usc -fcns}(\mathbb{R}^m; [0, 1]) \) if \( d^h(g', g) \to 0 \). Let \( N := \{1, 2, \ldots\} \). By Theorem 7.58 in [35], \( g' \) hypo-converges to \( g \) if and only if hypo \( g' \) set-converges\(^4\) to hypo \( g \). A well-known convenient characterization of hypo-convergence is given in [35, equation 7(9)], as follows.

\[ \text{PROPOSITION 2.1 (hypo-convergence). For } \{g, g'^\nu, \nu \in N\} \subset \text{usc -fcns}(\mathbb{R}^m; [0, 1]), \]
\[ g' \text{ hypo-converges to } g \text{ if and only if} \]
\[ (i) \text{ for all } \xi'^\nu \to \xi, \limsup_{\nu} g'^\nu(\xi'^\nu) \leq g(\xi); \]
\[ (ii) \text{ for all } \xi \text{ there exists a } \xi'^\nu \to \xi \text{ such that } \liminf_{\nu} g'^\nu(\xi'^\nu) \geq g(\xi). \]

In the following, we develop formulations of (AP) with \( Y = \text{usc -fcns}(\mathbb{R}^m; [0, 1]) \) and \( d_Y = d^h \); i.e., ambiguity sets are given in terms of distribution functions, and the analysis is carried out using the hypo-distance. We refer to such formulations as \textit{distribution-based}. Throughout, cd-fcns(\mathbb{R}^m) is viewed as a subset of the metric space \( \{ \text{usc -fcns}(\mathbb{R}^m; [0, 1]), d^h \} \).

3. Foundations for distribution functions. It is apparent that distribution-based formulations of stochastic ambiguity require us to develop the necessary mathematical tools for analyzing the inner maximization over usc functions. Since viewing distribution functions as a subset of usc functions might be beneficial in other situations too, we believe that the results in this section are of interest beyond the present context.

3.1. Weak convergence and other properties. Even though the space \( \text{usc -fcns}(\mathbb{R}^m; [0, 1]) \) is not linear, it is still meaningful to consider the weighted averages \( \lambda g + (1 - \lambda)g' \) for \( g, g' \in \text{usc -fcns}(\mathbb{R}^m; [0, 1]) \) and \( \lambda \in [0, 1] \), defined in a pointwise manner.\(^5\) The following definition is new.

\(^4\)We recall that the outer limit of a sequence of sets \( \{A'^\nu\}_{\nu \in N} \), denoted by Limsup \( A'^\nu \), is the collection of points \( a \) to which a subsequence of \( \{a'^\nu \in A'^\nu, \nu \in N\} \) converges. The inner limit, denoted by Liminf \( A'^\nu \), is the point to which a sequence of \( \{a'^\nu \in A'^\nu, \nu \in N\} \) converges. If both limits exist and are identical to \( A \), we say that \( \{A'^\nu\}_{\nu \in N} \) set-converges to \( A \), and we write \( A'^\nu \to A \).

\(^5\)For \( \xi \in \mathbb{R}^m \), \( (\lambda g + (1 - \lambda)g')(\xi) = \lambda g(\xi) + (1 - \lambda)g'(\xi). \)
Definition 3.1 (convexity on usc-fcns(\(\mathbb{R}^m;[0,1]\))). A set \(S \subset \text{usc-fcns}(\mathbb{R}^m;[0,1])\) is convex if for every \(g, g' \in S\) and \(\lambda \in [0,1]\), \(\lambda g + (1 - \lambda)g' \in S\).

A function \(\psi: S \to \mathbb{R}\) is convex if \(S \subset \text{usc-fcns}(\mathbb{R}^m;[0,1])\) is a convex set and if for every \(g, g' \in S\) and \(\lambda \in [0,1]\), \(\psi(\lambda g + (1 - \lambda)g') \leq \lambda \psi(g) + (1 - \lambda)\psi(g')\).

We note that usc-fcns(\(\mathbb{R}^m;[0,1]\)) and cd-fcns(\(\mathbb{R}^m\)) are convex. Obviously, the definition has the usual implication that every local minimizer of a convex function \(\psi\) over a convex set \(S\) is a global minimizer.

We recall that probability measures \(P^n \in \mathcal{M}\) converge weakly to a probability measure \(P \in \mathcal{M}\) if and only if \(\liminf_{\nu} P^n(A) = P(A)\) for all closed sets \(A \subset \mathbb{R}^m\). This convergence takes place if and only if the corresponding distribution functions \(F^n\) and \(F\) converge pointwise at all points of continuity of \(F\). The distribution functions are then said to also converge weakly. The following result is implicit in [46, 45], where the development is more abstract dealing with semicontinuous measures on closed sets. Here, we provide for the first time an explicit statement for the present context and give a new simplified proof that relies only on Proposition 2.1.

Theorem 3.2 (equivalence of hypo-convergence and weak convergence). For distribution functions \(F^n, F \in \text{cd-fcns}(\mathbb{R}^m)\) we have that \(d^b(F^n, F) \to 0\) if and only if \(F^n\) converges weakly to \(F\).

Proof. Let \(P^n, P \in \mathcal{M}\) correspond to \(F^n, F\); i.e., \(F^n(\xi) = P^n(S_\xi)\) and \(F(\xi) = P(S_\xi)\) for all \(\xi \in \mathbb{R}^m\). First, suppose that \(P^n\) converges weakly to \(P\). We utilize Proposition 2.1. For any \(\xi \in \mathbb{R}^m\), \(S_\xi\) is closed and therefore \(\limsup_{\nu} F^n(\xi) = \limsup_{\nu} P^n(S_\xi) \leq P(S_\xi) = F(\xi)\). Let \(\xi^\nu \to \xi\) and \(\varepsilon > 0\) be arbitrary. By the continuity from above of \(F\), there exists a \(\zeta \in \mathbb{R}^m\), with \(\zeta_i > 0\) for all \(i = 1, \ldots, m\), such that \(F(\xi + \zeta) \leq F(\xi) + \varepsilon\). Moreover, there exists a \(\bar{\nu} \in \mathbb{N}\) such that \(\xi^{\bar{\nu}} - \xi + \zeta\) for all \(\nu \geq \bar{\nu}\). Consequently, for all \(\nu \geq \bar{\nu}\), we have by the monotonicity of \(F^n\) that \(F^n(\xi^{\bar{\nu}}) \leq F^n(\xi + \zeta)\). This implies that \(\limsup_{\nu} F^n(\xi^{\nu}) \leq \limsup_{\nu} F^n(\xi + \zeta) \leq F(\xi) + \varepsilon\). Since \(\varepsilon\) is arbitrary, part (i) of Proposition 2.1 holds. Next, let \(\xi \in \mathbb{R}^m\) be arbitrary. Pick \(\zeta_i > 0\) for all \(i = 1, \ldots, m\). Then, there exists an open set \(A\) with \(S_\xi \subset A \subset S_{\xi + \zeta}\). Since \(A^c\) (the complement of \(A\)) is closed, \(\limsup P^n(A^c) \leq P(A^c)\), and therefore \(\liminf P^n(A) \geq P(A)\). Monotonicity implies that

\[
(2) \quad \liminf F^n(\xi + \zeta) = \liminf P^n(S_{\xi + \zeta}) \geq \liminf P^n(A) \geq P(A) \geq P(S_\xi) = F(\xi).
\]

Now, let \(\varepsilon > 0\), \(\nu_0 = 0\), \(\zeta^\mu \to 0\), with \(\zeta_i^\mu > 0\) for all \(i = 1, \ldots, m\) and \(\mu \in \mathbb{N}\). For all \(\mu\) there exists by (2) a \(\nu_\mu > \nu_{\mu - 1}\) such that, for all \(\nu \geq \nu_\mu\), \(F^n(\xi + \zeta^\mu) \geq F(\xi) - \varepsilon\). Construct the sequence \(\{\xi^{\nu}\}_{\nu \in \mathbb{N}}\) as follows. For \(\nu_1 \leq \nu \leq \nu_2\), set \(\xi^\nu = \xi + \zeta^\nu\). For \(k = 2, 3, \ldots, \), set \(\xi^\nu = \xi + \zeta^k\) for \(\nu_k < \nu \leq \nu_{k+1}\). By construction, \(F^n(\xi^{\nu}) \geq F(\xi) - \varepsilon\). Moreover, \(\xi^{\nu} \to \xi\), and part (ii) of Proposition 2.1 holds.

Second, we consider the converse and suppose that \(d^b(F^n, F) \to 0\). We directly obtain from part (i) of Proposition 2.1 that \(\limsup_{\nu} F^n(\xi) \leq F(\xi)\) for all \(\xi \in \mathbb{R}^m\). Let \(\xi \in \mathbb{R}^m\) be an arbitrary point at which \(F\) is continuous. We next establish that \(\liminf F^n(\xi) \geq F(\xi)\). Let \(\varepsilon > 0\). By the continuity of \(F\) at \(\xi\) there exists a \(\delta > 0\) such that \(F(\xi) \geq F(\xi - \delta)\) for all \(\xi \in \mathbb{R}^m\) satisfying \(\|\xi - \xi\| \leq \delta\). Let \(\zeta\) be one such point with \(\zeta_i < \xi_i\) for all \(i = 1, 2, \ldots, m\). Part (ii) of Proposition 2.1 implies that there exists a sequence \(\zeta^{\nu} \to \zeta\) such that \(\liminf F^n(\zeta^{\nu}) \geq F(\zeta)\). For this sequence, there exists a \(\bar{\nu} \in \mathbb{N}\) such that \(\zeta^{\nu} \leq \xi\) for all \(\nu \geq \bar{\nu}\). The monotonicity of \(F^n\) gives that \(F^n(\xi) \geq F^n(\zeta^{\nu})\) for all \(\nu \geq \bar{\nu}\). Combining these results, we obtain that

\[
\liminf F^n(\xi) \geq \liminf F^n(\zeta^{\nu}) \geq F(\zeta) \geq F(\xi) - \varepsilon.
\]
Since $\varepsilon$ is arbitrary, we have that $F^\nu(\xi) \to F(\xi)$ for $\xi$, a point of continuity of $F$. \hfill $\square$

An immediate consequence of this theorem is the fact that empirical distribution functions under independent sampling hypo-converge almost surely to the distribution function from which the samples are drawn; see, for example, [12, Theorem 11.4.1]. More complex examples of hypo-convergence of distribution functions are given below.

It is well known that the weak limit of a sequence of distribution functions might not be a distribution function. We recall that a subset $S \subset \text{cd-fcns}(\mathbb{R}^m)$ is tight if for all $\varepsilon > 0$ there exists a rectangle $A$ such that $\Delta AF \geq 1 - \varepsilon$ for all $F \in S$.

**Proposition 3.3 (convergence to distribution function).** If the sequence $\{F^\nu\}_{\nu \in \mathbb{N}} \subset \text{cd-fcns}(\mathbb{R}^m)$ is tight, then the following hold:

(i) There exists a subsequence $\{F^{\nu_k}\}_{k \in \mathbb{N}}$ and a function $F \in \text{cd-fcns}(\mathbb{R}^m)$ such that $d^h(F^{\nu_k}, F) \to 0$ as $k \to \infty$.

(ii) If $d^h(F^\nu, F) \to 0$ for some $F : \mathbb{R}^m \to \mathbb{R}$, then $F \in \text{cd-fcns}(\mathbb{R}^m)$.

**Proof.** In view of Theorem 3.2, the proof follows by standard results in probability theory; see, for example, [6, Theorem 29.3] and its corollary. \hfill $\square$

Generally, tightness is closely related to compactness as is well known from Prokhorov's theorem. Next we state for completeness the connection in the present context, but omit the straightforward proof. First, however, we establish some notation. We denote by $\text{cl} A$ the closure of a subset $A$ of a topological space. In any metric space $(S, d)$, we let the ball $B(u, r) := \{u' \in S : d(u, u') \leq r\}$. Moreover, $\mathcal{B} := B(0, 1)$ and $r\mathcal{B} := B(0, r)$ whenever the metric space contains a point 0. For $\mathbb{R}^k$, we use the usual Euclidean norm if not explicitly stated otherwise. We have already encountered an exception in the case of $\mathbb{R}^m \times \mathbb{R}$, which was given the norm $\| \cdot \|_S$ with balls $S(\xi, r)$. Still, we let $r\mathcal{S} := S(0, r)$.

**Proposition 3.4 (compactness, tightness).** For $S \subset \text{cd-fcns}(\mathbb{R}^m)$, we have that

(i) if $S$ is compact, then $S$ is tight;

(ii) if $S$ is tight, then $\text{cl} S$ is compact and contained in $\text{cd-fcns}(\mathbb{R}^m)$.

Although a ball $B(F, r) \subset \text{usc-fcns}(\mathbb{R}^m; [0, 1])$, with $F \in \text{cd-fcns}(\mathbb{R}^m)$, is compact, the subset $B(F, r) \cap \text{cd-fcns}(\mathbb{R}^m)$ is neither closed nor tight unless $r = 0$. This is easily seen in the following manner. Let $g : \mathbb{R}^m \to [0, 1]$ be defined such that $g(\xi) = \max\{0, F(\xi) - r\}$ and $r > 0$. Clearly, $d^h(g, F) \leq r$. Let $A \subset \mathbb{R}^m$ be such that $\Delta_A F \geq 1 - r$. Set $c > 0$ such that $A \subset \{\xi \in \mathbb{R}^m : \xi \leq c1\}$. Construct the distribution functions $F^\nu : \mathbb{R}^m \to [0, 1]$ such that $F^\nu(\xi) = 1$ if $c\nu 1 \leq \xi$ and $F^\nu(\xi) = g(\xi)$ otherwise. Since $F(\xi) \geq 1 - r$ for $\xi$ with $F^\nu(\xi) = 1$, we have that $|F^\nu(\xi) - F(\xi)| \leq r$ for all $\xi$ and thus $d^h(F^\nu, F) \leq r$. However, $\{F^\nu\}$ is not tight and does not tend to a distribution function. The only balls of $(\text{usc-fcns}(\mathbb{R}^m; [0, 1]), d^h)$ contained in $\text{cd-fcns}(\mathbb{R}^m)$ are those with zero radius, i.e., $\mathcal{B}(F, 0)$. We observe that a setup centered on the metric space $(\text{cd-fcns}(\mathbb{R}^m), d^h)$, instead of $(\text{usc-fcns}(\mathbb{R}^m; [0, 1]), d^h)$, is possible but has the disadvantage that the space is not complete and therefore is not Polish.

**Theorem 3.5 (estimate of the hypo-distance).** For $F, G \in \text{cd-fcns}(\mathbb{R}^m)$ and $\rho \in [1, \infty]$,

$$\eta(\rho)e^{-\rho} \leq d^h(F, G) \leq e^{-\rho} + (1 - e^{-\rho})\bar{\eta}(\rho) \leq \bar{\eta}(\infty),$$

\footnote{We use the notation $1 = (1, 1, \ldots, 1) \in \mathbb{R}^m$.}
where

\[ \bar{\eta}_m(\rho) = \inf \{ \eta \geq 0 : G(\xi + \eta \mathbf{1}/\sqrt{m}) + \eta \geq F(\xi), F(\xi + \eta \mathbf{1}/\sqrt{m}) + \eta \geq G(\xi) \ \forall \xi \in 2\rho B \} \]

and \[ \bar{\eta}_m(\rho) = \inf \{ \eta \geq 0 : G(\xi + \eta \mathbf{1}) + \eta \geq F(\xi), F(\xi + \eta \mathbf{1}) + \eta \geq G(\xi) \ \forall \xi \in \rho B \}. \]

**Proof.** The argumentation relies on quantities and associated notation introduced later and is postponed to the appendix.

We note that these estimates are obviously related but not identical to the Levy metric in the one-dimensional case \((m = 1)\). Since the hypo-distance is not scale invariant, it might be advantageous to shift and scale distribution functions such that \(\|\xi\|_2 \leq 1\) for “typical” points \(\xi \in \mathbb{R}^m\). This tends to make the comparison of horizontal/vertical distances, implicit in Theorem 3.5, more informative. For example, if \(a \in \mathbb{R}\), \(F(\xi) = \xi\) for \(\xi \in [0,1]\) (uniform distribution), and \(F = G(\cdot + a)\), then \(\bar{\eta}_m(\cdot) = a/2\), which is noninformative if \(a \geq 2\), as actually \(d^\rho(F,G) \leq 1\) always.

**Example 4: Hypo-distance for exponential distribution.** Theorem 3.5 provides the following bound on the hypo-distance between the distribution functions \(F\) and \(G\) given by \(F(\xi) = 0\) for \(\xi < 0\) and \(F(x) = 1 - \exp(-\lambda \xi)\) for \(\xi \geq 0\) (the exponential density) and \(G(\xi) = 0\) for \(\xi < 0\) and \(G(\xi) = 1\) for \(\xi \geq 0\). In this case, \(\bar{\eta}_m(\cdot)\) and \(\bar{\eta}_m(\cdot)\) are identical and equal to the Levy metric between \(F\) and \(G\) for any \(\rho \geq 1\). In fact, the value does not change for \(\rho \geq 1\). One can then find that \(\bar{\eta}_m(\cdot)\) is the root of \(\exp(-\lambda \eta) - \eta = 0\). For \(\lambda = 1, 10, 100,\) and \(1000\), the roots are 0.5671, 0.1746, 0.0339, and 0.0052, respectively. Thus, since these are independent of \(\rho \geq 1\), the upper bounds on the hypo-distances are simply these roots. The lower bound is scaled with \(\exp(-1)\).

**3.2. Ambiguity sets in distribution-based formulations.** The most interesting ambiguity sets inevitably arise in specific applications. Here, we simply illustrate some possibilities in distribution-based formulations.

**Example 5: Stochastic dominance.** First-order stochastic dominance (see, for example, [11]) can be used to construct ambiguity sets. Given two reference distribution functions \(F_0(\cdot, x), G_0(\cdot, x)\in\text{cd-fcns(}\mathbb{R}^m\text{)})\), which might depend on \(x \in \mathbb{R}^n\), the (ambiguity) set

\[ S(x) = \{ F \in \text{cd-fcns(}\mathbb{R}^m\text{)} : F_0(\xi, x) \leq F(\xi) \leq G_0(\xi, x) \ \forall \xi \in \mathbb{R}^m \} \]

specifies the consideration of all random vectors that are dominated in the usual first-order sense by a random vector with distribution function \(G_0(\cdot, x)\) and also by a random vector with distribution function \(F_0(\cdot, x)\) in the reoriented sense obtained by switching from the traditional focus on profit/gain to cost/loss; see [30]. For \(\xi \in \mathbb{R}^m\) and \(F' \in S(x) \rightarrow F\), part (i) of Proposition 2.1 implies that \(F_0(\xi, x) \leq F(\xi)\). Part (ii) of Proposition 2.1 and the fact that \(G_0(\cdot, x)\) is usc give that, for some \(\xi^* \rightarrow \xi\),

\[ G_0(\xi, x) \geq \limsup G_0(\xi^*, x) \geq \liminf F'(\xi^*) \geq F(\xi). \]

Thus, in view of the tightness of \(S(x)\) and Proposition 3.4, \(S(x)\) is compact. It follows directly from the definition that the set is also convex.

**Example 6: Quantile-type ambiguity.** Concentrating on a particular probability level \(\alpha(x) \in (0, 1)\) and a reference \(F_0(\cdot, x) \in \text{cd-fcns(}\mathbb{R}^m\text{)})\), which both might depend
on $x \in \mathbb{R}^n$, the set

$$S(x) = \{ F \in \text{cd-fcns}(\mathbb{R}^m) : \text{lev}_{\geq \alpha(x)} F_0(\cdot, x) \subset \text{lev}_{\geq \alpha(x)} F \},$$

with $\text{lev}_{\geq \beta} G := \{ \xi \in \mathbb{R}^m : G(\xi) \geq \beta \}$,
is a requirement that relates to $\alpha(x)$-efficient points [28]. In the case of $m = 1$, $F \in S(x)$ if and only if the $\alpha(x)$-quantile of $F$ is no larger than that of $F_0(\cdot, x)$. If $F^\nu \in S(x) \to F$, then $\text{lev}_{\geq \alpha(x)} F_0(\cdot, x) \subset \text{lev}_{\geq \alpha(x)} F$ by part (i) of Proposition 2.1 and $F \in S(x)$. Thus, the intersection of $S(x)$ with a closed subset of cd-fcns($\mathbb{R}^m$) is closed.

The previous two examples illustrate situations naturally suited to being handled using distribution-based formulations. Since weak convergence fails to ensure convergence of moments, the considerations of moments require additional restrictions. Before stating ambiguity sets in terms of moments, a common approach as seen in convergence of moments, the integral. For part (ii), let $u(F) = \int \psi(\xi)dF(\xi)$ is well-defined. Moreover, the following hold:

(i) $u$ is convex, provided that $S$ is a convex set. In fact, for $\lambda \in [0,1]$ and $F,G \in S$

$$u(\lambda F + (1-\lambda)G) = \lambda u(F) + (1-\lambda)u(G).$$

(ii) $u$ is lower semicontinuous (lsc) at $F \in S$ whenever the sets of discontinuity of $\psi$ and $F$ fail to intersect and $\psi \geq \beta \in \mathbb{R}$.

(iii) $u$ is continuous at $F \in S$ whenever the sets of discontinuity of $\psi$ and $F$ fail to intersect and, for every $F^\nu \in S \to F$,

$$\lim_{\gamma \to \infty} \sup_{\nu \in \mathbb{N}} \int_{\{\xi : |\psi(\xi)| \geq \gamma\}} |\psi(\xi)|dF^\nu(\xi) = 0.$$
development covers (AP) with ambiguity sets in a general metric space \((Y,d_Y)\), but we often provide details for specific formulations of stochastic ambiguity such as those that are distribution-based.

Associated with (AP), we consider a family of approximate problems

\[(AP)^\nu: \min_{x \in \mathcal{C}^\nu} \sup_{y \in D^\nu(x)} \varphi^\nu(x,y),\]

where the bifunction \(\varphi^\nu \in \text{bfcns}(\mathbb{R}^n,Y)\) specifies the cost and \(\mathcal{C}^\nu \subset \mathbb{R}^n, D^\nu: C^\nu \to Y\) specify \(\text{dom} \varphi^\nu = \{(x,y) \in \mathbb{R}^n \times Y : x \in C^\nu, y \in D^\nu(x)\}\). Throughout, we assume that the domain of \(\varphi^\nu\) is described in this manner by \(C^\nu\) and \(D^\nu\).

The approximate problems \(\{(AP)^\nu\}_{\nu \in \mathbb{N}}\) represent a multitude of situations arising in applications that demand “approximations” for computational and/or modeling reasons. We permit approximation in all components of (AP): the values of the cost function and its domain such as the region of feasible decisions and ambiguity set. Lopsided convergence, as defined next, deals with a notion of approximation of the bifunction \(\varphi\) defining (AP) by the bifunction \(\varphi^\nu\) specifying (AP)^\nu. It ensures convergence of minsup-points and minsup-values of (AP)^\nu to those of (AP), as we see in this section.

### 4.1. Definitions and consequences

In this subsection, we recall definitions and essential facts about lopsided convergence pertaining to the present context; see [41] for details and proofs. We start with the definition of lopsided convergence, which in addition to its “basic” form comes in two strengthened forms, referred to as ancillary-tight lop-convergence and tight lop-convergence.

**Definition 4.1 (lopsided convergence).** Let \(\{\varphi,\varphi^\nu, \nu \in \mathbb{N}\} \subset \text{bfcns}(\mathbb{R}^n,Y)\). The bifunctions \(\varphi^\nu\) converge lopsidedly, or lop-converge, to \(\varphi\), written \(\varphi^\nu \xrightarrow{\text{lop}} \varphi\), when the following hold:

(i) \(\forall x^\nu \in C^\nu \to x \in C, y \in D(x), \exists y^\nu \in D^\nu(x^\nu) \to y\) with \(\liminf \varphi^\nu(x^\nu,y^\nu) \geq \varphi(x,y)\), and \(\forall x^\nu \in C^\nu \to x \notin C, \exists y^\nu \in D^\nu(x^\nu)\) with \(\varphi^\nu(x^\nu,y^\nu) \to \infty\);

(ii) \(\forall x \in C, \exists x^\nu \in C^\nu \to x\) such that \(\forall y^\nu \in D^\nu(x^\nu) \to y \in Y, \limsup \varphi^\nu(x^\nu,y^\nu) \leq \varphi(x,y)\) if \(y \in D(x)\) and \(\varphi^\nu(x^\nu,y^\nu) \to -\infty\) otherwise.

The convergence is ancillary-tight if, in addition, for every \(\varepsilon > 0\) and \(x^\nu \to x\) selected in (ii), there exists a compact set \(B_x \subset Y\) and an integer \(\nu_x\) such that

\[
\sup_{y \in D^\nu(x^\nu) \cap B_x} \varphi^\nu(x^\nu,y) \geq \sup_{y \in D^\nu(x^\nu)} \varphi^\nu(x^\nu,y) - \varepsilon \quad \forall \nu \geq \nu_x.
\]

The convergence is tight if, in addition to all the above, for any \(\varepsilon > 0\) there exists a compact set \(A_x \subset \mathbb{R}^n\) and an integer \(\nu_x\) such that

\[
\inf_{x \in C^\nu \cap A_x} \sup_{y \in D^\nu(x)} \varphi^\nu(x,y) \leq \inf_{x \in C^\nu} \sup_{y \in D^\nu(x)} \varphi^\nu(x,y) + \varepsilon \quad \forall \nu \geq \nu_x.
\]

The requirements of ancillary-tight and tight lop-convergence can be viewed as relaxed “uniform” compactness assumptions on \(D^\nu(x)\) and \(C^\nu\). Obviously, ancillary-tightness is satisfied if all \(D^\nu(x^\nu)\) are contained in a compact set. Tightness is ensured if, in addition, all \(C^\nu\) are contained in a compact set. In both cases, many other situations without compactness will also satisfy the requirements.

A useful sufficient (but not necessary) condition for lop-convergence is stated in [41, Theorem 3.2], as given next.
Proposition 4.2 (sufficiency for lopsided convergence). Given \( \{\varphi, \varphi^\nu, \nu \in \mathbb{N}\} \subset \text{bfcns}(\mathbb{R}^n, Y) \), if the sets \( C^\nu = C \) are closed, \( D^\nu(x^\nu) \to D(x) \) whenever \( x^\nu \in C \to x \in C \), and \( \varphi^\nu(x^\nu, y^\nu) \to \varphi(x, y) \) whenever \( (x^\nu, y^\nu) \in \text{dom} \varphi^\nu \to (x, y) \in \text{dom} \varphi \), then \( \varphi^\nu \xrightarrow{\text{lop}} \varphi \).

The latter two conditions are referred to as continuous convergence.

Next, we turn to the implications of lopsided convergence and recall that

\[
\minsup \varphi := \inf_{x \in C} \sup_{y \in D(x)} \varphi(x, y)
\]

is the minsup-value of \( \varphi \), and denote by

\[
\varepsilon\text{-argminsup} \varphi := \left\{ x \in C : \sup_{y \in D(x)} \varphi(x, y) \leq \minsup \varphi + \varepsilon \right\}
\]

the set of \( \varepsilon \)-optimal solutions of (AP), \( \varepsilon \geq 0 \). If \( \varepsilon = 0 \), the set consists of the minsup-points of the bifunction \( \varphi \). Similar notation is adopted for \( \varphi^\nu \). The sup-projection of a bifunction \( \varphi \in \text{bfcns}(\mathbb{R}^n, Y) \), denoted by \( h \), is the real-valued function given by

\[
h(x) := \sup_{y \in D(x)} \varphi(x, y) \quad \text{whenever} \quad x \in C \quad \text{and} \quad \sup_{y \in D(x)} \varphi(x, y) < \infty.
\]

If \( \text{dom} h \) is empty, then \( \varphi \) has no sup-projection. When this happens, (AP) is in some sense "infeasible," as there is no \( x \in C \) that returns a finite value of the function to be minimized. This case is best treated separately, and we often exclude it below.

Proposition 4.3 (consequences of lop-convergence). Suppose that \( \{\varphi, \varphi^\nu, \nu \in \mathbb{N}\} \subset \text{bfcns}(\mathbb{R}^n, Y) \) have sup-projections, \( \varphi^\nu \xrightarrow{\text{lop}} \varphi \), and \( \{x^\nu \in \text{argminsup} \varphi^\nu, \nu \in \mathbb{N}\} \) exists. Then, the following hold:

(i) If \( \{x^\nu\}_{\nu \in \mathbb{N}} \) has a cluster point, then \( \liminf(\minsup \varphi^\nu) \geq \minsup \varphi \).

(ii) If lop-convergence is ancillary-tight, then \( \limsup(\minsup \varphi^\nu) \leq \minsup \varphi \) and

\[
\forall \{\varepsilon^\nu \searrow 0, \nu \in \mathbb{N}\}, \quad \text{Limsup}(\varepsilon^\nu\text{-argminsup} \varphi^\nu) \subset \text{argminsup} \varphi.
\]

(iii) If lop-convergence is ancillary-tight and \( \bar{x} \) is a cluster point of \( \{x^\nu\}_{\nu \in \mathbb{N}} \), i.e., the limit of a subsequence \( \{x^{\nu_k}\}_{k \in \mathbb{N}} \), then \( \lim_k(\minsup \varphi^{\nu_k}) = \minsup \varphi \).

(iv) If lop-convergence is tight, then \( \minsup \varphi^\nu \to \minsup \varphi \in \mathbb{R} \) and there exists \( \{\varepsilon^\nu \downarrow 0, \nu \in \mathbb{N}\} \) such that \( \varepsilon^\nu\text{-argminsup} \varphi^\nu \to \text{argminsup} \varphi \) even if the assumption that \( \{x^\nu \in \text{argminsup} \varphi^\nu, \nu \in \mathbb{N}\} \) exists is dropped.

It is apparent from this proposition, which is a collection of results from [41], that lopsided convergence becomes a central property in the study of approximations of (AP) and their solutions.

The infimum of a real-valued lsc function with domain contained in a compact set is attained. Hence, when \( \sup_{y \in D(.)} \varphi(\cdot, y) \) is lsc and \( C \) is contained in a compact set, then there exists a minsup-point of \( \varphi \). We next state an existence result for minsup-points that relaxes the compactness requirement. The result compiles Theorem 4.9 and Proposition 5.1 in [41].

Proposition 4.4 (existence of minsup-point). Suppose that \( \{\varphi, \varphi^\nu, \nu \in \mathbb{N}\} \subset \text{bfcns}(\mathbb{R}^n, Y) \) have sup-projections, \( \varphi^\nu \xrightarrow{\text{lop}} \varphi \) ancillary-tightly, and there are compact points that relaxes the compactness requirement. The result compiles Theorem 4.9 and Proposition 5.1 in [41].
sets \( \{ B^\nu \subset \mathbb{R}^n, \nu \in \mathbb{N} \} \) such that \( C^\nu \subset B^\nu \). Moreover, suppose that, for each \( \nu \), \( D^\nu \) is inner semicontinuous\(^8\) and \( \varphi^\nu \) is lsc.

Then, for all \( \nu \) there exists a minsup-point \( x^\nu \) of \( \varphi^\nu \), and every cluster point of \( \{ x^\nu \}_{\nu \in \mathbb{N}} \) is a minsup-point of \( \varphi \).

If for all \( \nu \) we have that \( D^\nu \subset Y \) (i.e., the mapping is constant), then the condition of lower semicontinuity of \( \varphi^\nu \) can be replaced by lower semicontinuity of \( \varphi^\nu (\cdot, y) \) for all \( y \in D^\nu \).

Proposition 4.4 does not ensure the existence of a cluster point of \( \{ x^\nu \}_{\nu \in \mathbb{N}} \). Still, the proposition provides an approach for establishing the existence of a minsup-point of (AP) without requiring compactness of \( C \) and \( D(x) \) for \( x \in C \). One can first construct a sequence \( \{ \varphi^\nu \}_{\nu \in \mathbb{N}} \), with the required properties, that lop-converges to \( \varphi \) ancillary-tightly and, second, prove that \( \{ x^\nu \}_{\nu \in \mathbb{N}} \) has a cluster point.

4.2. Lop-convergence in optimization under stochastic ambiguity. Relying on the previous subsection, we next develop a series of specific results under various assumptions about (AP) and \( (AP)^\nu \). However, we start with an algorithm that solves (AP) by computing approximate solutions of the approximations \( (AP)^\nu \).

**Algorithm for (AP).**

**Step 0:** Select \( \varepsilon^\nu \downarrow 0 \) and set \( \nu = 1 \).

**Step 1:** Compute \( x^\nu \in \varepsilon^\nu \text{-argminsup} \varphi^\nu \).

**Step 2:** Replace \( \nu \) by \( \nu + 1 \), and go to Step 1.

**Theorem 4.5 (algorithmic convergence).** Suppose that the sup-projections of the bifunctions \( \{ \varphi, \varphi^\nu, \nu \in \mathbb{N} \} \subset \text{bfcns}(\mathbb{R}^n, Y) \) are well-defined, that these bifunctions lop-converge ancillary-tightly to \( \varphi \) with argminsup \( \varphi^\nu \neq \emptyset \) for all \( \nu \in \mathbb{N} \), and that the Algorithm for (AP) has generated the sequence \( \{ x^\nu \}_{\nu \in \mathbb{N}} \).

If \( \bar{x} \) is the limit of a subsequence \( \{ x^{\nu_k} \}_{k \in \mathbb{N}} \) of \( \{ x^\nu \}_{\nu \in \mathbb{N}} \), then \( \lim_{\nu \to \infty} (\text{minsup} \varphi^{\nu_k}) = \text{minsup} \varphi \) and \( \bar{x} \in \text{argminsup} \varphi \).

**Proof.** The result follows from a direct application of Proposition 4.3. \(\square\)

We next give a series of conditions under which (ancillary-tight) lop-convergence holds and concentrate mostly on distribution-based formulations, i.e., the case \( Y = \text{usc-fcns}(\mathbb{R}^m; [0, 1]) \). For such formulations, it is obvious in view of Proposition 3.4 that \( \varphi^\nu \text{ lop } \varphi \) implies ancillary-tightly lop-convergence when \( \{ D^\nu (x^\nu) \}_{\nu \in \mathbb{N}} \) is tight for every sequence \( x^\nu \to x \) selected in Definition 4.1(ii).

The next example shows the useful result that if the cost function in (AP) is approximated in some “continuous” way, then ambiguity sets constructed from a “robust band” around empirical distribution functions lead to lopsided convergence. This provides a justification for the strategy “robustify” an empirical distribution function that is feared to deviate substantially from the actual distribution function.

**Example 8: Lop-convergence under sampling.** In a distribution-based formulation, suppose that \( \varphi \in \text{bfcns}(\mathbb{R}^n, \text{usc-fcns}(\mathbb{R}^m; [0, 1])) \) has \( \text{dom} \varphi = C \times D \), where \( C \subset \mathbb{R}^n \) is closed and \( D = \{ F_0 \} \subset \text{cd-fcns}(\mathbb{R}^m) \). That is, suppose that the actual problem has no ambiguity. Let \( \{ \xi^\nu \}_{\nu \in \mathbb{N}} \) be a sequence of independent random \( m \)-vectors, distributed as \( F_0 \) and defined on a probability space \( (\Omega, \mathcal{A}, \mu) \), and let \( \{ \varepsilon^\nu : C \to [0, \infty), \nu \in \mathbb{N} \} \) be such that for any \( x^\nu \in C \to x \in C, \varepsilon^\nu (x^\nu) \downarrow 0 \). The approximations

\(\footnote{A mapping \( S : C \rightrightarrows Y \) is inner semicontinuous if for all \( x^\nu \in C \to x \in C, \liminf S(x^\nu) \supset S(x) \).} \)
\( \varphi^* \in \text{bfcns}(\mathbb{R}^n, \text{usc-fcns}(\mathbb{R}^m; [0, 1])) \) have \( C^* = C \) and, for \( x \in C \),

\[
D^*(x) = \{ F \in \text{cd-fcns}(\mathbb{R}^m) : \mathcal{d}^h(F, F^*_{\omega}) \leq \varepsilon^*(x) \},
\]

where \( F^*_{\omega}(\xi) = (1/\nu) \sum_{j=1}^{\nu} I(\xi^j(\omega) \leq \xi) \), \( \xi \in \mathbb{R}^m \), are the corresponding empirical distribution functions, with \( I(\xi^j \leq \xi) = 1 \) if \( \xi^j \leq \xi \) and 0 otherwise. If \( \varphi^* \) converges continuously to \( \varphi \), then \( \varphi^* \) lop-converges ancillary-tightly to \( \varphi \) for \( \mu \)-almost every \( \omega \in \Omega \).

This fact can be established using Proposition 4.2. Obviously, \( F^*_{\omega} \to F_0 \) for \( \mu \)-almost every \( \omega \in \Omega \). Let \( \omega \) be such that this convergence takes place. We start by showing continuous convergence of \( D^* \) to \( D \). Let \( x^* \in C \to x \in C \). First, we establish that Limsup \( D^*(x^*) \subset \{ F_0 \} \). If \( F \in \text{Limsup} D^*(x^*) \), then there exists \( F^k \in D^{\nu_k}(x^{\nu_k}) \to F \). Let \( \delta > 0 \). There exists a \( k \in \mathbb{N} \) such that \( \varepsilon^{\nu_k}(x^{\nu_k}) \leq \delta/3 \), \( \mathcal{d}^h(F, F^k) \leq \delta/3 \), and \( \mathcal{d}^h(F_0, F^{\nu_k}) \leq \delta/3 \) for all \( k \geq k \). For such \( k \),

\[
\mathcal{d}^h(F_0, F) \leq \mathcal{d}^h(F, F^k) + \mathcal{d}^h(F^k, F^{\nu_k}) + \mathcal{d}^h(F_0, F^{\nu_k}) \leq \delta.
\]

Since \( \delta > 0 \) is arbitrary, we conclude that \( F = F_0 \). Second, we establish that \( F_0 \in \text{Liminf} D^*(x^*) \). Set \( F^{\nu} = F^{\nu_k} \). Then, \( F^{\nu} \in D^*(x^*) \) for all \( \nu \) and \( F^{\nu} \to F_0 \). Thus, \( D^* \) converges continuously to \( \{ F_0 \} \). This fact, together with the continuous convergence of \( \varphi^* \) to \( \varphi \), establishes that \( \varphi^* \) lop-converges to \( \varphi \) almost surely. The lop-convergence is in fact ancillary-tight because for \( \varepsilon > 0 \) and \( x^* \to x \) we can define \( B_\varepsilon = \mathbb{B}(F_0, \varepsilon) \), which is compact, and then find that for sufficiently large \( \nu \), \( \mathcal{d}^h(F^{\nu}, F_0) \leq \varepsilon/2 \) and \( \varepsilon^*(x^*) \leq \varepsilon/2 \).

In addition to empirical distribution functions, there are numerous other paths to constructing arbitrarily accurate approximations of a distribution function as well as other functions in the sense of \( \mathcal{d}^h \).

**Example 9: Epi-spline approximations of ambiguity sets.** Suppose that \( \varphi \in \text{bfcns}(\mathbb{R}^n, \text{usc-fcns}(\mathbb{R}^m; [0, 1])) \) has domain \( C \times D \), with \( C \subset \mathbb{R}^n \) closed and the ambiguity set \( D \) solid, i.e., \( D \) equals the closure of its interior. We consider an epi-spline approximation of this ambiguity set. Suppose that \( \{ S^\nu \}_{\nu=1}^{\infty} \) is an infinite refinement of \( \mathbb{R}^m \) in the sense of [42]; i.e., each partition \( S^\nu = \{ S_k^\nu \}_{k=1}^{N^\nu} \) has \( S_k^\nu \subset \mathbb{R}^m \) open, \( \cup_{k=1}^{N^\nu} S_k^\nu = \mathbb{R}^m \), and \( S_0^\nu \cap S_{N^\nu}^\nu = \emptyset \), \( k \neq l \). Moreover, \( S_k^\nu \) gradually becomes “smaller” as precisely stated in [42]. A simple example of an infinite refinement of \( \mathbb{R} \) is to take \( N^\nu = 2\nu + 2 \), \( S_1^\nu = (-\infty, -\sqrt{\nu}) \), \( S_k^\nu = ((k-\nu-2)/\sqrt{\nu}, (k-\nu-1)/\sqrt{\nu}) \) for \( k = 2, 3, \ldots, 2\nu+1 \), and \( S_{2\nu+2}^\nu = (\sqrt{\nu}, \infty) \). An epi-spline \( s \) of order \( p \in \mathbb{N} \) on partition \( S^\nu \) is a real-valued function on \( \mathbb{R}^m \) that on each \( S_k^\nu \) is polynomial with total degree \( p \) and that satisfies \( \liminf s(\xi^l) = s(\xi) \) for all \( \xi^l \to \xi \). Let e-spl \( (S^\nu) \) be the collection of all such epi-splines. An approximation of \( \varphi \) in terms of epi-splines could then take the following form. Let \( \varphi^* \in \text{bfcns}(\mathbb{R}^n, \text{usc-fcns}(\mathbb{R}^m; [0, 1])) \) have \( C^* = C \) and

\[
D^*(x) = \{ g \in \text{usc-fcns}(\mathbb{R}^m; [0, 1]) : \mathcal{d}^h(g, g^*) \leq \varepsilon^*(x) \text{ for some } g^* \in D \},
\]

where \( \{ \varepsilon^* : C \to [0, \infty) \} \) is such that, for any \( x^* \in C \to x \in C \), \( \varepsilon^*(x^*) \to 0 \). In addition, suppose that \( \varphi^* \) converges continuously to \( \varphi \).

We show that \( \varphi^* \) converges to \( \varphi \), using Proposition 4.2. First, we establish Limsup \( D^*(x^*) \subset D \) when \( x^* \in C \to x \in C \). Set \( g \in \text{Limsup} D^*(x^*) \). Then there exists \( g^k \in D^{\nu_k}(x^{\nu_k}) \to g \). We also have \( g^k \in D \) with \( \mathcal{d}^h(g^k, g^*) \leq \varepsilon^{\nu_k}(x^{\nu_k}) \). Thus, \( g^k \to g \) and, by the closedness of \( D \), \( g \in D \), which establishes the assertion. Second, we show that Liminf \( D^*(x^*) \subset D \) when \( x^* \to x \). Set \( g \in D \). Since \( D \) is the closure of its interior,
there exists a sequence \( \{g^\mu\}_{\mu \in \mathbb{N}} \) in the interior of \( D \) such that \( \delta^h(g^\mu, g) \to 0 \). For each \( \mu \), the fact that \( g^\mu \) belongs to an open subset of \( D \) implies that there exists a sequence \( \{g^{\nu^\mu}\}_{\nu^\mu \in \mathbb{N}} \subset D \), with \(-\bar{g}^{\nu^\mu} \in e\text{-spl}(\mathcal{S}^\nu)\), such that \( \bar{g}^{\nu^\mu} \to g^\mu \) as \( \nu \to \infty \); see [42, Theorem 3.5] and Proposition 4.2 guarantees that \( \bar{g}^{\nu^\mu} \to g^\mu \) for all \( \nu \geq \nu^\mu \). Consequently, we can construct a sequence \( \{\nu^\mu\}_{\mu \in \mathbb{N}} \) such that \( \nu^\mu \geq \nu^\mu \geq \max\{\nu_0, \nu^\mu + 1\} \) and \( \delta^h(\bar{g}^{\nu^\mu}, g^\nu) \leq 1/\mu \) for all \( \nu \geq \nu^\mu \) and \( \mu \). For arbitrary \( \delta > 0 \) there exists therefore an integer \( \mu_0 \geq 2/\delta \) such that, for all \( \nu \geq \nu^\mu \) and \( \mu > \mu_0 \), \( \delta(g^\mu, g) \leq \delta/2 \) and \( \delta^h(\bar{g}^{\nu^\mu}, g^\nu) \leq \delta/2 \). We construct the sequence \( \{f^\nu\}_{\nu \in \mathbb{N}} \) by setting \( f^\nu = \bar{g}^{\nu^\mu} \) with \( \mu \) satisfying \( \nu \in \{\nu^\mu - 1, \nu^\mu - 2, \ldots, \nu^\mu\} \). Then for every \( \nu > \nu^\mu \) and some \( \nu_\mu > \nu_0 \),

\[
\delta^h(f^\nu, g) = \delta(\bar{g}^{\nu^\mu}, g) \leq \delta^h(\bar{g}^{\nu^\mu}, g^\nu) + \delta^h(g^\nu, g) \leq \delta/2 + \delta/2 = \delta.
\]

Consequently, \( \delta^h(f^\nu, g) \to 0 \). Since \( f^\nu \in D^\nu(x^\nu) \supset D \) and therefore \( D^\nu(x^\nu) \to D \). This fact in conjunction with Proposition 4.2 ensures that \( \varphi^\nu \xrightarrow{\text{hyp}} \varphi \). A similar result holds in the case of restriction to distribution functions on \( \mathbb{R}^2 \) using only a finite number of parameters.

**Example 10: Approximation of stochastic dominance.** Suppose that the bifunctions \( (\varphi, \varphi^\nu, \nu \in \mathbb{N}) \in \text{bfcns}(\mathbb{R}^m, \text{usc-fcns}(\mathbb{R}^m; [0, 1])) \), with \( \varphi^\nu \) converging continuously to \( \varphi \), and their domains are specified as follows. The sets \( C \in C^\nu \) are closed, and for all \( x \in C \) suppose that \( \{F_0, G_0, F_0^\nu(x), G_0^\nu(x), \nu \in \mathbb{N}\} \subset \text{cd-fcns}(\mathbb{R}^m) \) are reference distribution functions, which might depend on \( x \in C \), and \( F_0^\nu, G_0^\nu \) converge continuously to \( F_0, G_0 \), respectively. The set-valued mappings associated with the bifunctions \( \varphi, \varphi^\nu \) are then specified by

\[
D(x) = \{F \in \text{cd-fcns}(\mathbb{R}^m) : F_0(\xi, x) \leq F(\xi) \leq G_0(\xi, x) \forall \xi \in \mathbb{R}^m\},
\]

\[
D^\nu(x) = \{F \in \text{cd-fcns}(\mathbb{R}^m) : F_0^\nu(\xi, x) \leq F(\xi) \leq G_0^\nu(\xi, x) \forall \xi \in \Xi^\nu\},
\]

where \( \Xi^\nu \subset \mathbb{R}^m \) set-converges to \( \mathbb{R}^m \). In this case, the stochastic dominance constraint is approximated in two ways: the reference distribution functions are approximated and the constraints are enforced on a smaller set, possibly consisting of a finite number of points. (Of course, since \( \mathbb{R}^m \) is separable, there are finite sets \( \Xi^\nu \) that set-converge to \( \mathbb{R}^m \).) Under the stated assumptions, \( \varphi^\nu \xrightarrow{\text{hyp}} \varphi \), provided that \( D^\nu \) converges continuously to \( D \), which is indeed the case, as we see next. Fix \( x^\nu \in C \to x \in C \). First, take \( F \in \text{Limsup} D^\nu(x^\nu) \). Then there exists \( F^k \in D^{x^\nu}(x^{x^\nu}) \to F \), and thus for all \( \xi \in \Xi^\nu, F_0^\nu(\xi, x^{x^\nu}) \leq F^k(\xi) \leq G_0^\nu(\xi, x^{x^\nu}) \). Let \( \xi \in \mathbb{R}^m \). There exists \( x^{x^\nu} \in \Xi^\nu \to \xi \) by the assumption about \( \Xi^\nu \). Since \( F^k \to F \),

\[
F(\xi) \geq \text{limsup}_k F^k(\xi^{x^\nu}) \geq \text{limsup}_k F_0^\nu(\xi^{x^\nu}, x^{x^\nu}) = F_0(\xi, x).
\]

Thus, the lower bound is satisfied, and we turn our attention to the upper bound. Let \( \varepsilon > 0 \). The continuous convergence of \( G^\nu_0 \) to \( G_0 \) ensures that there exist \( \delta > 0 \) and \( k_1 \in \mathbb{N} \) such that

\[
G^\nu_0(\eta, x^{x^\nu}) \leq G_0(\xi, x) + \varepsilon/2,
\]

provided that \( \|\eta - \xi\| \leq \delta \) and \( k \geq k_1 \).

Select \( \zeta \in \mathbb{R}^m \) such that \( 0 < \|\zeta - \zeta\| < \delta \) and the components \( \zeta_i > \zeta_i \) for all \( i = 1, \ldots, m \). Since \( \Xi^\nu \to \mathbb{R}^m \), there exist \( \zeta^k \in \Xi^\nu \to \zeta \) and \( k_2 \geq k_1 \) such that \( 0 < \|\zeta - \zeta^k\| < \delta \) and \( \zeta^k > \zeta_i \) for all \( i = 1, \ldots, m \) and \( k \geq k_2 \). The hypo-convergence

\[
G^\nu_0(\eta, x^{x^\nu}) \leq G_0(\xi, x) + \varepsilon/2,
\]

provided that \( \|\eta - \xi\| \leq \delta \) and \( k \geq k_1 \).
of $F^k$ to $F$ guarantees that there exists $\xi^k \to \xi$ such that $\liminf F^k(\xi^k) \geq F(\xi)$. Thus, there exists $k_3 \geq k_2$ such that $F^k(\xi^k) + \varepsilon/2 \geq F(\xi)$ and $\xi^k \leq \xi^k$ for all $k \geq k_3$. Collecting these facts and using the monotonicity of $F^k$, we obtain that for all $k \geq k_3$

$$F(\xi) \leq F^k(\xi^k) + \frac{\varepsilon}{2} \leq F^k(\xi^k) + \frac{\varepsilon}{2} \leq G^0_{\nu^k}(\xi^k, x^\nu^k) + \frac{\varepsilon}{2} \leq G_0(\xi, x) + \varepsilon.$$ 

Since $\varepsilon > 0$ is arbitrary, we also conclude that $F$ satisfies the upper bound. The tightness of $\{F^k\}_{k \in \mathbb{N}}$ implies that $F$ is a distribution function, and thus $F \in D(x)$ and $\limsup D^\nu(x^\nu) \subset D(x)$.

Second, take $F \in D(x)$ and construct $F^\nu : \mathbb{R}^m \to [0, 1]$, by setting

$$F^\nu(\xi) = \max\{F_0^\nu(\xi, x^\nu), \min[\nu(F(\xi), G_0(\xi, x^\nu))]\} \quad \text{for } \xi \in \mathbb{R}^m,$$

which is a distribution function, and $F^\nu \in D^\nu(x^\nu)$. We show that $F^\nu \to F$ by means of Proposition 2.1. For part (i) of that proposition, let $\nu^\nu \to \nu$. Then, by the continuous convergence of $F_0^\nu, G_0^\nu$ and the upper semicontinuity of $F$, $\limsup F^\nu(\nu^\nu) \leq \max\{F_0(\nu, x), \min[F(\xi), G_0(\nu, \xi, x)]\} = F(\nu)$.

For part (ii) of the proposition, let $\nu^\nu' = \nu$. Then, $\lim\inf F^\nu(\nu^\nu') = \max\{F_0(\nu, x), \min[F(\xi), G_0(\nu, \xi, x)]\} = F(\nu)$. Thus, $F^\nu \to F$, and therefore $D(\nu) \subset \liminf F^\nu(x^\nu)$. Consequently, $D^\nu(x^\nu) \to D(x)$ and $\varphi^\nu \top F \varphi$ by Proposition 4.2.

**Example 11: Approximation of quantile-type ambiguity.** Suppose that the bifunctions $\{\varphi, \varphi^\nu, \nu \in \mathbb{N}\} \subset \text{bfcns}(\mathbb{R}^m, \text{usc-fcns}(\mathbb{R}^m; [0, 1]))$ have domains characterized by the closed sets $C = C^\nu$ and the set-valued mappings

$$D(\nu) = \{F \in S_0 : \text{lev}_{\geq \alpha(\nu)} F_0(\cdot, x) \subset \text{lev}_{\geq \alpha(\nu)} F\},$$

$$D^\nu(\nu) = \{F \in S_0 : \text{lev}_{\geq \alpha^\nu(\nu)} F_0^\nu(\cdot, x) \subset \text{lev}_{\geq \alpha^\nu(\nu)} F\},$$

where $S_0 \subset \text{cd-fcns}(\mathbb{R}^m)$ is closed and convex, $\alpha, \alpha^\nu : C \to (0, 1)$, and $F_0^\nu(\cdot, x), F_0(\cdot, x) \in S_0$ for all $x \in C$. We make the following assumptions about the approximations: the bifunctions $\varphi^\nu$ converge continuously to $\varphi$, and for every $x^\nu \in C \to x \in C$ and $\gamma^\nu \to \gamma$, $\text{lev}_{\geq \alpha^\nu} F_0^\nu(\cdot, x^\nu) \to \text{lev}_{\gamma} F_0(\cdot, x)$ and $\alpha^\nu(x^\nu) \to \alpha(x)$. We observe that this “continuous convergence” of superlevel sets is stronger than $F_0^\nu(\cdot, x^\nu) \to F_0(\cdot, x)$.

To establish that $\varphi^\nu \top F \varphi$ it suffices to show the continuous convergence of $D^\nu$ to $D$; cf. Proposition 4.2. Let $x^\nu \in C \to x$. First, we prove that $\limsup D^\nu(x^\nu) \subset D(x)$. Take $F \in \limsup D^\nu(x^\nu)$. Then there exists $F^k \in D^\nu(x^\nu_k) \to F$ and

$$\text{lev}_{\geq \alpha^\nu(\nu_k)} F_0^\nu(\cdot, x^\nu_k) \subset \text{lev}_{\geq \alpha(\nu_k(x^\nu_k))} F^k.$$ 

We seek to establish that $\text{lev}_{\geq \alpha(\nu)} F_0(\cdot, x) \subset \text{lev}_{\geq \alpha(\nu)} F$. Hypo-convergence implies a certain “convergence” of superlevel sets [35, Proposition 7.7]: for any $\beta^k \in \mathbb{R} \to \beta \in \mathbb{R}$, $\limsup_k \text{lev}_{\geq \beta^k} F^k \subset \text{lev}_{\geq \beta} F$. Thus,

$$\text{lev}_{\geq \alpha(\nu)} F_0(\cdot, x) \subset \text{Liminf}_k \left(\text{lev}_{\geq \alpha^\nu(\nu_k)} F_0^\nu(\cdot, x^\nu_k)\right)$$

$$\subset \text{Limsup}_k \left(\text{lev}_{\geq \alpha^\nu(\nu_k)} F^k\right) \subset \text{lev}_{\geq \alpha(\nu)} F.$$

Second, we establish that $\text{Liminf } D^\nu(x^\nu) \supset D(x)$. Set $F \in D(x)$. We construct $F^\nu : \mathbb{R}^m \to [0, 1]$ by setting for $\xi \in \mathbb{R}^m$

$$F^\nu(\xi) = \max\{\alpha^\nu(x^\nu), F(\xi)\} \quad \text{if } F_0^\nu(\xi, x^\nu) \geq \alpha^\nu(x^\nu) \quad \text{and } F^\nu(\xi) = F(\xi) \quad \text{otherwise},$$

and
which indeed is a distribution function. The convexity of $S_0$ ensures that $F'' \in S_0$. If $\xi \in \mathbb{R}^m$ is such that $F'_0(\xi, x^\nu) \geq \alpha^\nu(x^\nu)$, then $F''(\xi) = \max\{\alpha^\nu(x^\nu), F(\xi)\} \geq \alpha^\nu(x^\nu)$ and

$$\text{lev}_{\geq \alpha^\nu(x^\nu)} F''(\cdot, x^\nu) \subset \text{lev}_{\geq \alpha^\nu(x^\nu)} F''.$$

Consequently, $F'' \in D''(x^\nu)$. We now show that $F'' \to F$, using Proposition 2.1. Let $\xi' \to \xi$. We aim to show that $\limsup F''(\xi') \leq F(\xi)$. Since $F$ is usc, the condition is automatic if $F'_0(\xi', x^\nu) \geq \alpha^\nu(x^\nu)$ holds only for a finite number of $\nu$. If there is a subsequence $\{\xi'_{k}\}_{k \in \mathbb{N}}$, such that $F'_0(\xi'_{k}, x^\nu) \geq \alpha^\nu(x^\nu)$, then from the continuous convergence assumptions, $F'_0(\xi, x) \geq \alpha(x)$). Thus, $F(\xi) \geq \alpha(x)$ also. Moreover, because $F''(\xi') \leq \max\{\alpha^\nu(x^\nu), F(\xi')\}$, we have that $\limsup F''(\xi') \leq \max\{\alpha(\nu), F(\xi)\} = F(\xi)$, and part (i) of Proposition 2.1 is shown. For part (ii), select $\xi'' = \xi$. Since $F''(\xi) \geq \liminf F''(\xi'') \geq \limsup F''(\xi')$, we also have that $\liminf F''(\xi) = F(\xi)$. Consequently, $F'' \to F$ and $\liminf D''(x^\nu) \supset D(x)$. We therefore have that $\varphi'' \cdot \varphi$.

**Example 12: Approximations of moments.** We consider bifunctions $\{\varphi, \varphi', \nu \in \mathbb{N}\}$, with $\varphi''$ converging continuously to $\varphi$, that have domains given, in part, by moment-type restrictions. Specifically, suppose that $C = C''$ are closed, $\psi: \mathbb{R}^m \to \mathbb{R}^p$ is continuous, and $S_0 \subset \text{cd-fcns}(\mathbb{R}^m)$ is a closed set on which the uniform integrability property holds, i.e., $\lim_{\gamma \to \infty} \sup_{F \in S_0} \int_{\{\xi: \psi(\xi) \geq \gamma\}} \|\psi(\xi)\|_\infty dF(\xi) = 0$. The set-valued mappings that complete the specification of the domains of $\varphi''$ and $\varphi$ are given for $x \in C$ as

$$D(x) = \{F \in S_0 : u(F) \in T(x) \subset \mathbb{R}^p\},$$

$$D'(x) = \{F \in S_0 : u(F) \in T'(x) \subset \mathbb{R}^p\},$$

where $u(F) := \int \psi(\xi) dF(\xi)$. The set-valued mappings $T'': C \to \mathbb{R}^p$ converge continuously to $T: C \to \mathbb{R}^p$. Under these assumptions, it suffices to show that $D''$ converges continuously to $D$ for $\varphi'' \cdot \varphi$ to hold: cf. Proposition 4.2.

Let $x^\nu \in C \to x \in C$. First, we show that $\limsup D''(x^\nu) \subset D(x)$. Suppose that $F \in \limsup D''(x^\nu) \subset D(x)$. Then there exists $F^k \in D''(x^\nu) \to F$. In view of Proposition 3.6, $u$ is continuous on $S_0$. This fact, together with $T''(x^\nu) \to T(x)$, implies that $u(F) \in T(x)$. Then, $F \in D(x)$, and the claim is established. Second, we need to ensure that $\liminf D''(x^\nu) \supset D(x)$. Let $F \in D(x)$. We need to construct $F'' \in D''(x^\nu) \to F$. That is, for some given $\mu'' \in T''(x^\nu) \to u(F)$, $F''$ needs to be selected such that $u(F'') = \mu''$ and $F'' \to F$. Here, additional assumptions need to be brought in, as this is not possible in general. We consider two possibilities:

(a) Suppose that the additional condition is $T(x) \subset T''(x^\nu)$. This situation addresses applications where $T(x)$ is a singleton giving the (true) moments and $T''(x^\nu)$ is a confidence region for the moments constructed using samples; see [10] for a similar situation. At least with high probability, the confidence region will contain the moment values and will shrink as the sample size ($\nu$) grows. Regardless of the application, we here can set $F'' = F$. Since $u(F) \in T(x) \subset T''(x^\nu)$, we certainly have $F'' \in D''(x^\nu) \to F$.

(b) Suppose the additional conditions are that $S_0$ is convex, $T(x) = [\alpha(x), \beta(x)] \subset \mathbb{R}$, $T''(x^\nu) = [\alpha''(x^\nu), \beta''(x^\nu)] \subset \mathbb{R}$, with $\beta''(x^\nu) \geq \beta(x)$, and there is an $F^0 \in S_0$ with $u(F^0) > \beta^1$. (A similar argument as the one that follows deals with the case when $T''(x^\nu)$ “approaches” from below.) Since $T''(x^\nu) \to T(x)$, there exists $\mu'' \in$
By construction, $\lambda' \in [0, 1] \to 0$. Since $S_0$ is convex, $F' \in S_0$, and by part (i) of Proposition 3.6, $u(F') = \mu'$. Thus, $F' = D'(x')$. It now remains only to show that $F' \to F$. Let $\xi' \to \xi$. Then, $\limsup F'(\xi') \leq \limsup \lambda' + F(\xi') \leq F(\xi)$ because $\lambda' \to 0$. Since $F'$ also converges pointwise to $F$, $F' \to F$.

In both (a) and (b), we therefore have that Liminf $D'(x') \supset D(x)$. Thus, $D'(x') \to D(x)$ and $\varphi' \xrightarrow{l.o.s.} \varphi$ in view of Proposition 4.2.

**Example 13: Approximation in superquantile-risk minimization.** We recall the situation and notation in Example 2 and a superquantile-risk minimization problem

$$
\min_{x \in C} \sup_{q \in D(x)} \varphi(x, q),
$$

where $\varphi \in \text{bfcns}(\mathbb{R}^n, \mathcal{L}^\infty(\mathbb{R}^m))$ has domain specified by the closed set $C \subset \mathbb{R}^n$ and

$$D(x) = \left\{ q \in \mathcal{L}^\infty(\mathbb{R}^m) : 0 \leq q(\xi) \leq \frac{1}{1 - \alpha(x)}, \text{P}_0\text{-a.e. } \xi \in \mathbb{R}^m, \int q(\xi)dP_0(\xi) = 1 \right\}
$$

with $\alpha : C \to [0, 1]$ now depending on $x$. Given $\psi(x, \cdot) \in \mathcal{L}^1(\mathbb{R}^m)$ for all $x \in C$, the bifunction takes the values

$$
\varphi(x, q) = \int \psi(x, \xi)q(\xi)dP_0(\xi) \quad \text{for } x \in C \text{ and } q \in D(x).
$$

Since $\alpha(x)$ can be interpreted as the level of risk averseness [30, 32], we here examine a situation with decision-dependent risk averseness, a setting rarely considered. This is an example of an ambiguity set that is not given in terms of distribution functions. Here, $Y = \mathcal{L}^\infty(\mathbb{R}^m)$, which is a metric space under $d_Y(\cdot, \cdot) = \|\cdot - \cdot\|_\infty$ and the usual consideration of equivalence classes, where $\|q\|_\infty = \inf\{\eta \geq 0 : |q(\xi)| \leq \eta \text{ for } P_0\text{-a.e. } \xi \in \mathbb{R}^m\}$. Suppose that $\alpha$ and $\psi$ are approximated by $\alpha' : C \to [0, 1]$ and $\psi : C \times \mathbb{R}^m \to \mathbb{R}$, where $\psi'(x, \cdot) \in \mathcal{L}^1(\mathbb{R}^m)$ for all $x \in C$. The approximate bifunctions $\varphi' \in \text{bfcns}(\mathbb{R}^n, \mathcal{L}^1(\mathbb{R}^m))$ are then defined as in the case of $\varphi$, but with $\alpha$ and $\psi$ replaced by $\alpha'$ and $\psi'$, respectively. Moreover, for all $x' \in C \to x \in C$, $\alpha'(x') \to \alpha(x)$ and for some $\Xi \subset \mathbb{R}^m$ with $P_0(\Xi) = 1$

$$
\sup_{\xi \in \Xi} |\psi'(x', \xi) - \psi(x, \xi)| \to 0 \quad \text{whenever } x' \to x \in C.
$$

We show that $\varphi' \xrightarrow{\text{l.o.s.}} \varphi$ by establishing that $\varphi'$ converges continuously to $\varphi$ and $D'$ to $D$. We start with the latter and let $x' \in C \to x \in C$. For $q \in \text{Limsup} D'(x')$ there exists $q^k \in D'(x') \to q$. Since $\|q^k\|_\infty \leq 1/(1 - \alpha(x'))$ and $\alpha'(x') \to \alpha(x)$, $\|q^k\|_\infty \leq 1/(1 - \alpha(x))$. Moreover, $\int q(\xi)dP_0(\xi) = \int q(\xi) - q^k(\xi)dP_0(\xi) + 1 \to 1$ because $\|q^k - q\|_\infty \to 0$. Thus, $q \in D(x)$ and $\text{Limsup} D'(x') \subset D(x)$. We next establish that Liminf $D'(x') \supset D(x)$. Given $q \in D(x)$, we construct $\tilde{q}''(\xi) = \min\{q(\xi), 1/(1 - \alpha''(x'))\}$. Certainly, $\|\tilde{q}''\|_\infty \leq 1/(1 - \alpha''(x'))$. Let the integral $\int \tilde{q}'' : = \int \max\{0, q(\xi) - 1/(1 - \alpha''(x'))\}dP_0(\xi)$. If $\tilde{q}'' = 0$, then set $q'' = \tilde{q}''$. Otherwise, $q'' = \tilde{q}'' + w''$, where $w''$ is constructed as follows. Since $\int [1/(1 - \alpha''(x'))] - \tilde{q}''(\xi)dP_0(\xi) = 1/(1 - \alpha''(x')) - 1 + \tilde{q}'' > \tilde{q}''$ in this case, we have that there exists a measurable

$$T'(x') \to u(F).$$

The additional conditions permit us to select $\mu' \geq u(F)$. Set

$$F' = \lambda'F^0 + (1 - \lambda')F, \quad \text{where } \lambda' = \frac{\mu' - u(F)}{u(F^0) - u(F)}.$$

Therefore, $F' \to F$. Let $\xi' \to \xi$. Then, $\limsup F'(\xi') \leq \limsup \lambda' + F(\xi') \leq F(\xi)$ because $\lambda' \to 0$. Since $F'$ also converges pointwise to $F$, $F' \to F$.

In both (a) and (b), we therefore have that Liminf $D'(x') \supset D(x)$. Thus, $D'(x') \to D(x)$ and $\varphi' \xrightarrow{\text{l.o.s.}} \varphi$ in view of Proposition 4.2.
functions \( w^\nu : \mathbb{R}^m \to [0, \infty) \), with \( w^\nu(\xi) \leq 1/(1 - \alpha^\nu(x^\nu)) - \tilde{q}^\nu(\xi) \) for \( \xi \in \mathbb{R}^m \), such that \( \int w^\nu(\xi)dP_0(\xi) = \nu^\nu \). Thus, \( q^\nu \in D^\nu(x^\nu) \). Let \( S^\nu = \{ \xi \in \mathbb{R}^m : w^\nu(\xi) > 0 \} \). The functions \( w^\nu \) can be selected such that \( \{ P_0(S^\nu) \}_{\nu \in \mathbb{N}} \) is bounded away from zero. Hence, there exists a constant \( c < \infty \) such that \( \| w^\nu \|_\infty \leq c \nu^\nu \) for all \( \nu \). Since \( \nu^\nu \to 0 \), \( \| w^\nu \|_\infty \to 0 \) and also \( q^\nu = q^0 + w^\nu \to q \). We have shown that \( \text{Liminf} D^\nu(x^\nu) \supset D(x) \) and therefore \( D^\nu(x^\nu) \to D(x) \). We next consider the continuous convergence of \( \varphi^\nu \) to \( \varphi \). Let \( (x^\nu, q^\nu) \in \text{dom} \varphi^\nu \to (x, q) \in \text{dom} \varphi \). In view of the assumption about \( \varphi^\nu \) and \( \varphi \),

\[
|\varphi^\nu(x^\nu, q^\nu) - \varphi(x, q)| \leq \int |\psi^\nu(x^\nu, \xi)q^\nu(\xi) - \psi(x, \xi)q(\xi)|dP_0(\xi) \to 0.
\]

Consequently, \( \varphi^\nu \xrightarrow{\text{lop}} \varphi \), by Proposition 4.2. The approximation of \( \alpha \) by \( \alpha^\nu \) addresses, for example, a situation where the level of risk averseness is unknown and we are interested in “stability” in optimal solutions under changes in risk averseness. Since lop-convergence holds, which implies convergence of the solutions (see Proposition 4.3), we have a certain continuity property in solutions of superquantile-risk minimization problems as functions of the risk averseness.

**Example 14:** Robust optimization and generalized semi-infinite programming. Suppose that \( \varphi \in \text{bfcns}(\mathbb{R}^n, Y) \) defines the problem

\[
\min_{x \in C} \sup_{y \in D(x)} \varphi(x, y),
\]

where \( C \subset \mathbb{R}^n \) and \( D_0 \subset Y \) are closed and where the set-valued mapping \( D : C \to Y \) further specifies the domain of \( \varphi \) by having

\[
D(x) = \{ y \in D_0 \subset Y : g(x, y) \leq 0 \}, \quad \text{with } g : C \times D_0 \to \mathbb{R}^p \text{ continuous.}
\]

In terms of a continuous function \( \varphi_0 : C \times D_0 \to \mathbb{R} \), the value of \( \varphi \) is specified as

\[
\varphi(x, y) = \varphi_0(x, y) \quad \text{for } x \in C, \ y \in D(x).
\]

This problem can be solved by considering the approximate problems

\[
\min_{x \in C} \sup_{y \in D_0} \varphi_0(x, y) - t^\nu g(x, y)^+,
\]

where \( v^+ = \max\{0, v_1, v_2, \ldots, v_p\} \) for any vector \( v = (v_1, \ldots, v_p) \) and \( t^\nu \to \infty \) are positive penalties; see [40, 39] for algorithms along these lines and [51] for a general treatment. This is another example of an ambiguity set that is not distribution-based.

The approximate problem is fully characterized by the bifunctions \( \varphi^\nu \in \text{bfcns}(\mathbb{R}^n, Y) \), with

\[
\varphi^\nu(x, y) = \varphi_0(x, y) - t^\nu g(x, y)^+ \quad \text{for } x \in C^\nu = C, \ y \in D^\nu(x) = D_0.
\]

Suppose that \( D \) is a continuous set-valued mapping on \( C \). Lop-convergence of \( \varphi^\nu \) to \( \varphi \) can then be proven by means of Definition 4.1 as follows. We note that this is a case where the ambiguity set \( D(x) \) is not approximated directly, yet lop-convergence can still be established. First we consider part (i) of Definition 4.1 and let \( x^\nu \in C \to x \in C \) and \( y \in D(x) \). Select \( y^\nu \in D(x^\nu) \subset D^\nu(x^\nu) \to y \). Such a sequence obviously exists in view of the continuity assumption about \( D \). Thus, \( \varphi^\nu(x^\nu, y^\nu) = \)
\( \varphi_0(x', y') - t'g(x', y')^+ = \varphi_0(x', y') \to \varphi_0(x, y) = \varphi(x, y) \), which establishes the first requirement. For part (ii) of Definition 4.1, let \( h \) be continuous. Since \( \varphi \) is continuous and \( g \) is continuous, it follows that \( \varphi_0(x', y') - t'g(x', y')^+ \to \varphi_0(x, y') - t'g(x', y')^+ \to -\infty \) because \( t' \to \infty \), which establishes lopsided convergence. Thus, through Proposition 4.3, this provides a justification for algorithms based on the solution of the approximate problem.

5. Quantification of lop-convergence and solution estimates. We established in [41] that the lop-distance quantifies lopsided convergence. Leveraging this fundamental result, we here give bounds for this distance and use them to estimate rates of convergence and solution errors for optimization problems with stochastic objectives.

5.1. Lop-distance. We recall that the sup-projection of \( \varphi \in \text{bfcns}(\mathbb{R}^n, Y) \) is denoted by \( h \); similarly let \( h^\nu \) be the sup-projection of the approximate bifunction \( \varphi^\nu \in \text{bfcns}(\mathbb{R}^n, Y) \); i.e., \( h^\nu \) is the real-valued function given by

\[
h^\nu(x) := \sup_{y \in D^\nu(x)} \varphi^\nu(x, y) \quad \text{whenever } x \in C^\nu \text{ and } \sup_{y \in D^\nu(x)} \varphi^\nu(x, y) \leq \infty.
\]

If \( \text{dom} h^\nu \) is empty, then \( \varphi^\nu \) has no sup-projection. It is obvious that the minsup-points of \( \varphi^\nu \) are identical to the minimizers of \( h(h^\nu) \). Thus, sup-projections will be central to the following development. In fact, the key quantity is the epi-distance between two sup-projections, as defined next.

Although our primary interest is in the epi-distance between sup-projections \( h \) and \( h^\nu \), which are defined on subsets of \( \mathbb{R}^n \), no complications arise from stating some definitions and results for a general metric space \((X, d_X)\). We reap the benefits of this approach in some proofs when applying the results to the inner problem on the space of 

\[
\text{lsc-fcns}(X; \mathbb{R}) := \{ f : C \to \mathbb{R} : \text{for some nonempty } C \subset X \text{ with closed epi } g \}. 
\]

Let \( \bar{x} \in X \) be arbitrary and fixed throughout. If \( X \) is Euclidean, then we always set \( \bar{x} = 0 \). The following results hold for any \( \bar{x} \), but conclusions might be more practically useful if problem instances are appropriately rescaled and shifted such that function values are comparable to \( d_X(\bar{x}, x) \) for “typical” \( x \in X \). We refer to \( \bar{x} \) as the centroid of \( X \).

For \( f, f' \in \text{lsc-fcns}(X; \mathbb{R}) \), let the epi-distance

\[
d_f^\rho(f, f') := \int_0^\infty d_f^\rho(f, f') e^{-\rho} d\rho,
\]

where the \( \rho \)-epi-distance, \( \rho \geq 0 \), is given by

\[
d_f^\rho(f, f') := \max \{|\text{dist}(\bar{x}, \text{epi } f) - \text{dist}(\bar{x}, \text{epi } f')| : \bar{x} \in S_f^\rho\},
\]

\[
d_f^\rho(f, f') := \max \{|\text{dist}(\bar{x}, \text{epi } f) - \text{dist}(\bar{x}, \text{epi } f')| : \bar{x} \in S_f^\rho\},
\]

where the \( \rho \)-epi-distance, \( \rho \geq 0 \), is given by

\[
d_f^\rho(f, f') := \max \{|\text{dist}(\bar{x}, \text{epi } f) - \text{dist}(\bar{x}, \text{epi } f')| : \bar{x} \in S_f^\rho\},
\]

\[
d_f^\rho(f, f') := \max \{|\text{dist}(\bar{x}, \text{epi } f) - \text{dist}(\bar{x}, \text{epi } f')| : \bar{x} \in S_f^\rho\},
\]

where the \( \rho \)-epi-distance, \( \rho \geq 0 \), is given by
with $S := \mathbb{B}_{\rho} \times [-\rho, \rho]$, $\mathbb{B}_{\rho} := \overline{B}(\hat{x}, \rho)$, and
\[
\text{dist}((x, x_0), S) := \inf \{ \max[|d_X(x, z)|, |z_0 - z_0|] : (z, z_0) \in S \}
\]
if $S \subset X \times \mathbb{R}$ is nonempty and $\text{dist}(\hat{x}, 0) = \infty$. When $X$ is infinite-dimensional, then $d^\rho$ is called the *Attouch–Wets distance*, as it metrizes the Attouch–Wets topology. The parallel with Wets distance, $f$, obvious and, for $f$, and $\text{dist}(\bar{\nu}, \nu) = 0$, whenever $d^\rho(f', f) \to 0$. The converse holds if $X$, $d_X$ is proper, i.e., its closed balls are compact as in the case $X = \mathbb{R}^n$ and $d_X = \| \cdot \|_2$; see, for example, [38].

With this background, we are ready to state the definition of lop-distance between bifunctions. Since we rely on the epi-distance between the corresponding sup-projections, we limit the scope to bifunctions with sup-projections in $\text{lsc-fcns}(\mathbb{R}^n; \mathbb{R})$. This is not a strong assumption because, for example, if $\varphi \in \text{bfcns}(\mathbb{R}^n, Y)$ is lsc as a function of both variables and the set-valued mapping $D$ is inner semicontinuous, then the sup-projection of $\varphi$ is lsc when it exists. Alternatively, if $D \subset Y$, i.e., the set-valued mapping is constant on $C$, then it suffices to have $\varphi(\cdot, y)$ lsc for all $y \in D$; see Proposition 5.1 in [41].

Our main motivation for defining the lop-distance is to apply it in the study of minsup-points of bifunctions. Thus, informally, we would like to say that two bifunctions are close if their minsup-points and minsup-values are close, or equivalently, that the optimal solutions and optimal values of the corresponding sup-projections are close. As we see below, this is indeed the case if the sup-projections corresponding to the bifunctions are close in the sense of the epi-distance. Consequently, we settled on the following definition [41]: The *lop-distance* between $\varphi, \varphi' \in \text{bfcns}(\mathbb{R}^n, Y)$, with sup-projections $h$ and $h'$ in $\text{lsc-fcns}(\mathbb{R}^n; \mathbb{R})$, is
\[
d^\rho(\varphi, \varphi') := d^\rho(h, h').
\]

Theorem 5.4 in [41] establishes that if $\varphi'$ lop-converges ancillary-tightly to $\varphi$ and $\{ \varphi, \varphi', \nu \in \mathbb{N} \}$ have sup-projections in $\text{lsc-fcns}(\mathbb{R}^n; \mathbb{R})$, then $d^\rho(\varphi', \varphi) \to 0$. The converse also holds in some sense after passing to equivalence classes; see [41] for details. Consequently, the lop-distance provides a central tool in estimating the distance between bifunctions and their minsup-points and minsup-values. Next, for the first time we set out to estimate the lop-distance.

### 5.2. Estimates of lop-distance.

For $f, f' \in \text{lsc-fcns}(X; \mathbb{R})$, we define the auxiliary quantity
\[
\hat{d}^\rho(f, f') := \max \left\{ e \left( \text{epi } f \cap S, \text{epi } f' \right), e \left( \text{epi } f' \cap S, \text{epi } f \right) \right\},
\]
where the *excess* of a set $S$ over a set $S'$ is given by
\[
e(S, S') := \sup \{ \text{dist}(z, S') : z \in S \}
\]
if $S, S'$ are nonempty, $e(S, S') = \infty$ if $S$ is nonempty and $S'$ is empty, and $e(S, S') = 0$ otherwise. When $X = \mathbb{R}^n$, the auxiliary quantity has the following alternative expression:
\[
\hat{d}^\rho(f, f') = \inf \left\{ \eta \geq 0 : \text{epi } f \cap \rho S \subset \text{epi } f' + \eta S, \text{epi } f' \cap \rho S \subset \text{epi } f + \eta S \right\},
\]
where $S$ is the unit ball at the origin of $\mathbb{R}^n \times \mathbb{R}$ under the norm $\| \cdot \|_2$; see (1). This auxiliary quantity is usually more accessible than $d^\rho$, as seen below. Applied to the sup-projections of the bifunctions of interest, it provides a key estimate.
PROPOSITION 5.1 (estimates of lop-distance). For $\varphi, \varphi' \in \text{bfcns}(\mathbb{R}^n, Y)$, with sup-projections $h, h' \in \text{isc-fcns}(\mathbb{R}^n; \mathbb{R})$, we have for any $\rho \geq 0$,

$$(1 - e^{-\rho})|d - d'| + e^{-\rho}d_x^e(h, h') \leq d^e(\varphi, \varphi') \leq (1 - e^{-\rho})d_x^e(h, h') + e^{-\rho}(\max\{d, d'\} + \rho + 1),$$

where $d = \text{dist}(0, \text{epi } h)$, $d' = \text{dist}(0, \text{epi } h')$, and $\bar{\rho} \geq 2\rho + \max\{d, d'\}$.

Proof. This result can be deduced from a more general result in [38, Proposition 3.1] and also parallels, though with a different norm, [35, Exercise 7.60], which omits a direct proof. We provide a new proof simplifying arguments of the former source and take inspiration from the proofs of [35, Lemmas 4.34, 4.41]. Clearly,

$$d^e(h, h') = \int_0^\rho d_x^e(h, h')e^{-\tau}d\tau + \int_\rho^\infty d_x^e(h, h')e^{-\tau}d\tau.$$

Since $d_x^e(h, h')$ is nondecreasing as $\tau$ increases, we have that

$$d_x^e(h, h') \int_0^\rho e^{-\tau}d\tau \leq \int_0^\rho d_x^e(h, h')e^{-\tau}d\tau \leq d_x^e(h, h') \int_0^\rho e^{-\tau}d\tau$$

and

$$d_x^e(h, h') \int_\rho^\infty e^{-\tau}d\tau \leq \int_\rho^\infty d_x^e(h, h')e^{-\tau}d\tau \leq \int_\rho^\infty [\max\{d, d'\} + \tau]e^{-\tau}d\tau,$$

where the last inequality follows from the fact that $d_x^e(h, h') \leq \max\{d, d'\} + \tau$, by the triangle inequality. Carrying out the integrations on the left- and right-hand sides, we obtain that

$$(1 - e^{-\rho})|d - d'| + e^{-\rho}d_x^e(h, h') \leq d^e(h, h') \leq (1 - e^{-\rho})d_x^e(h, h') + e^{-\rho}(\max\{d, d'\} + \rho + 1).$$

(3)

Next, we establish the relation between $d_x^e$ and $d_x^c$. Suppose that $C$ and $D$ are closed subsets of $\mathbb{R}^n \times \mathbb{R}$; that $\varepsilon > 0$, $\rho > 0$; and that $\rho' \geq 2\rho + \text{dist}(0, C)$. We first show that $\text{dist}(\cdot, D) \leq \text{dist}(\cdot, C) + \varepsilon$ on $\rho S$ implies that $C \cap \rho S \subset D + \varepsilon S$. The claim is trivial if $C$ is empty. For nonempty $C$, we have for every $\bar{x} \in C \cap \rho S$ that

$$\text{dist}(\bar{x}, D) \leq \varepsilon.$$

As $D$ is closed, we have that $C \cap \rho S \subset D + \varepsilon S$. Second, we establish that $C \cap \rho' S \subset D + \varepsilon S$ implies $\text{dist}(\cdot, D) \leq \text{dist}(\cdot, C) + \varepsilon$ on $\rho S$. For any $\bar{x} \in \mathbb{R}^{n+1}$,

$$\text{dist}(\bar{x}, C \cap \rho' S) \geq \text{dist}(\bar{x}, D + \varepsilon S) = \inf\{\|\bar{y} + \varepsilon \bar{z}\|_S : \bar{y} \in D, \bar{z} \in S\}$$

$$\geq \text{inf}\{\|\bar{y} - \bar{x}\|_S - \varepsilon\|\bar{z}\|_S : \bar{y} \in D, \bar{z} \in S\} = \text{dist}(\bar{x}, D) - \varepsilon.$$

Thus, $\text{dist}(\cdot, D) \leq \text{dist}(\cdot, C \cap \rho' S) + \varepsilon$ on $\mathbb{R}^{n+1}$. It remains to establish that $\text{dist}(\bar{x}, C \cap \rho' S) = \text{dist}(\bar{x}, C)$ when $\bar{x} \in \rho S$ and $\rho' \geq 2\rho + \text{dist}(0, C)$. So let $\bar{x} \in \rho S$ and $\bar{y} \in \text{argmin}_{\bar{y} \in C} \|\bar{x} - \bar{y}\|_S$, which exists since $C$ is closed. The implication is established if $\bar{y} \in \rho' S$. This is indeed the case because $\|\bar{y}\|_S \leq \|\bar{x}\|_S + \|\bar{y} - \bar{x}\|_S$, with $\|\bar{y} - \bar{x}\|_S = \text{dist}(\bar{x}, C) \leq \text{dist}(\bar{x}, 0) + \text{dist}(0, C)$. Consequently, $\|\bar{y}\|_S \leq 2\|\bar{x}\|_S + \text{dist}(0, C) \leq 2\rho + \text{dist}(0, C) \leq \rho'$. Applying these two implications, first with $C = \text{epi } h$ and $D = \text{epi } h'$ and second with $C = \text{epi } h'$ and $D = \text{epi } h$, we obtain that

$$d_x^c(h, h') \leq d_x^c(h, h') \leq d_x^c(h, h') \quad \text{for } \bar{\rho} \geq 2\rho + \max\{d, d'\}.$$
A bound on $\hat{d}_x^\rho$ is given next, which is tight when $X = \mathbb{R}^n$. For $f : C \subset X \to \mathbb{R}$, we adopt the notation $\text{lev}_{\leq \rho} f := \{ x \in C : f(x) \leq \rho \}$ and
\[
\inf f := \inf_{x \in C} f(x) \quad \text{and} \quad \text{argmin } f := \{ x \in C : f(x) \leq \inf f \}.
\]
Moreover, for $B \subset X$, $\inf_B f := \inf_{x \in B \cap C} f(x)$, which is interpreted as $\infty$ when $B \cap C = \emptyset$.

**Proposition 5.2** (bound for auxiliary quantity). For $f, f'' \in \text{lsc-fcns}(X; \mathbb{R})$ and $\rho \geq 0$, $\hat{d}_x^\rho(f, f'') \leq \eta_\rho(f, f'')$, where
\[
\eta_\rho(f, f'') := \inf \left\{ \eta \geq 0 : \inf_{\mathbb{B}(x, \eta)} f'' \leq \max\{f(x), -\rho\} + \eta \forall x \in \mathbb{B}_\rho \cap \text{lev}_{\leq \rho} f \right\}
\]
\[
\inf_{\mathbb{B}(x, \eta)} f \leq \max\{f''(x), -\rho\} + \eta \forall x \in \mathbb{B}_\rho \cap \text{lev}_{\leq \rho} f''\}
\]
If $X = \mathbb{R}^n$, then the relation holds with equality.

**Proof.** This result is given in [38, Proposition 3.2]. We provide a new proof when $X = \mathbb{R}^n$, which is much shorter than the proof of the more general result. It also serves to correct a flawed argument in the proof of [35, Proposition 7.61]. Since we are dealing with epi-graphs on both sides, $\eta \geq 0$ satisfies $\text{epi } f \cap \rho S \subset \text{epi } f'' + \eta S$ if and only if it satisfies
\[
\text{epi } f \cap \left( (\rho \mathbb{B} \cap \text{lev}_{\leq \rho} f) \times [-\rho, \infty) \right) \subset \text{epi } f'' + \eta (\mathbb{B} \times [-1, \infty)),
\]
where $\mathbb{B}$ is the unit ball at the origin of $\mathbb{R}^n$. This is also equivalent to
\[
\text{epi } \left( \max\{f, -\rho\} + \delta_{\rho \mathbb{B} \cap \text{lev}_{\leq \rho} f} \right) \subset \text{epi } f'' + \text{epi}(\delta_\mathbb{B} - \eta),
\]
where $\delta_S$ is the function on $\mathbb{R}^n$ that equals 0 on $S \subset \mathbb{R}^n$ and $\infty$ otherwise. In view of [35, Exercise 1.28], we observe that
\[
\text{epi } f'' + \text{epi}(\delta_\mathbb{B} - \eta) = \text{epi } f''_{\eta}, \quad \text{with } f''_{\eta}(x) = \inf_{\mathbb{B}(x, \eta)} f'' - \eta.
\]
Thus, $\text{epi } f \cap \rho S \subset \text{epi } f'' + \eta S$ holds if and only if $\inf_{\mathbb{B}(x, \eta)} f'' \leq \max\{f(x), -\rho\} + \eta$ for all $x \in \rho \mathbb{B} \cap \text{lev}_{\leq \rho} f$. The conclusion follows after a similar argument with the roles of $f$ and $f''$ reversed.

We state the implication of these results and definitions for estimates of minsup-points and minsup-values. Although these estimates could have been given in terms of the lop-distance, simplifications accrue from working directly with $\eta_\rho(h, h'')$. In view of Propositions 5.1 and 5.2, the difference is minor anyway. We state the result for general functions and specialize to bifunctions in a corollary. Let $\mathbb{R}_+ = [0, \infty)$.

**Proposition 5.3** (estimates of optimal values and solutions). Suppose that $f, f'' \in \text{lsc-fcns}(X; \mathbb{R})$ and $\rho \in \mathbb{R}_+$ are such that
\[
\rho \geq \inf f \geq -\rho \quad \text{and} \quad \text{argmin } f \cap \mathbb{B}_\rho \neq \emptyset,
\]
with a similar condition for $f''$. Then,
\[
|\inf f'' - \inf f| \leq \eta_\rho(f, f'').
\]
If, in addition, there exists an increasing and continuous function \( \psi_f : \mathbb{R}_+ \to \mathbb{R}_+ \), with \( \psi_f(0) = 0 \), such that

\[
    f(x) - \inf f \geq \psi_f(\text{dist}(x, \text{argmin } f)) \quad \forall x \in \mathbb{B}_{\rho'},
\]

where \( \rho' > \rho + \eta_{\rho}(f, f') \), then

\[
    e(\arg\min f' \cap \mathbb{B}_\rho, \arg\min f) \leq \eta_{\rho}(f, f') + \psi_{\rho}^{-1}(2\eta_{\rho}(f, f')).
\]

Proof. The proof is nearly identical to those of Theorems 4.1 and 4.2 in [38] and is omitted. The only difference is a slight change in assumption about the size of \( \rho \) in the first part, and a relaxation of the requirement on \( \psi_f \) in the second part.

It is easy to find examples where these upper bounds are strict. An understanding of the conditioning function \( \psi_f \) might stem from the simple observation that if \( f(x) = \|x\|_2 \), then we can select \( \psi_f(\tau) = \tau^2 \) and \( \psi_{\rho}^{-1}(\eta) = \sqrt{\eta} \). The previous theorem yields the following corollary.

**Corollary 5.4 (estimates of minsup-points and values).** Suppose that \( \varphi, \varphi' \in \text{bfcns}(\mathbb{R}^n, Y) \) have sup-projections \( h, h' \in \text{lsc-fcns}(\mathbb{R}^n; \mathbb{R}) \), \( \rho \in \mathbb{R}_+ \), and

\[
    \rho \geq \text{minsup } \varphi \geq -\rho \quad \text{and} \quad \text{argminsup } \varphi \cap \rho \mathbb{B} \neq \emptyset,
\]

with a similar requirement on \( \varphi' \). Then,

\[
    |\text{minsup } \varphi' - \text{minsup } \varphi| \leq \hat{d}_\rho^e(h, h').
\]

If, in addition, there exists an increasing and continuous function \( \psi_h : \mathbb{R}_+ \to \mathbb{R}_+ \), with \( \psi_h(0) = 0 \), such that

\[
    h(x) - \text{minsup } \varphi \geq \psi_h(\text{dist}(x, \text{argmin } \varphi)) \quad \forall x \in \rho' \mathbb{B},
\]

where \( \rho' > \rho + \hat{d}_\rho^e(h, h') \), then

\[
    (\arg\minsup \varphi' \cap \rho \mathbb{B}) \subset \text{argminsup } \varphi + \left(\hat{d}_\rho^e(h, h') + \psi_h^{-1}(2\hat{d}_\rho^e(h, h'))\right) \mathbb{B}.
\]

Corollary 5.4 is significant, as it directly ties the distances between minsup-points and minsup-values of two bifunctions to the corresponding auxiliary quantity \( \hat{d}_\rho^e \) and, through Proposition 5.1, the lop-distance. In fact, the lop-distance was constructed with this goal in mind.

A useful estimate of \( \hat{d}_\rho^e \) is given next. We say that a function \( f : X \to \mathbb{R} \) is Lipschitz continuous with modulus \( \kappa : \mathbb{R}_+ \to \mathbb{R}_+ \) if

\[
    |f(x) - f(x')| \leq \kappa(\rho)d_X(x, x') \quad \forall x, x' \in \mathbb{B}_\rho.
\]

A component towards an estimate is the distance between two nonempty closed subsets \( S, S' \subset X \) given by

\[
    \hat{d}_\rho(S, S') := \max \left\{ e(S \cap \mathbb{B}_\rho, S'), e(S' \cap \mathbb{B}_\rho, S) \right\}, \quad \rho \geq 0.
\]

This quantity is closely related to \( \hat{d}_\rho^e \) but deals with arbitrary nonempty closed sets, not only epi-graphs. The next proposition is a slight simplification of [38, Proposition 3.3], and the proof is omitted.
Proposition 5.5. Suppose that $f, f' \in \text{ls-c-fcns}(X; \mathbb{R})$ have closed domains $C$ and $C'$, respectively, and in terms of Lipschitz continuous $f_0, f'_0 : X \to \mathbb{R}$, with common modulus $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$, are given by

$$f = f_0 \text{ on } C \quad \text{and} \quad f' = f'_0 \text{ on } C'.$$

Then, for $\rho \in \mathbb{R}_+$ and $\rho' > \rho + \hat{d}_\rho(C, C')$,

$$\hat{d}_\rho^\prime(f, f') \leq \eta_{\rho}(f, f') \leq \sup_{x \in \mathbb{R}_+} |f_0(x) - f'_0(x)| + \max\{1, \kappa(\rho')\} \hat{d}_\rho(C, C').$$

If $(X, d_X)$ is proper, then $\rho' = \rho + \hat{d}_\rho(C, C')$ is permissible and $\hat{d}_\rho^\prime(f, f') = \eta_{\rho}(f, f')$.

We say that $\varphi_0 : \mathbb{R}^n \times Y \to \mathbb{R}$ is marginally Lipschitz continuous with moduli $\kappa_1 : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ and $\kappa_2 : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ if, for every $\rho, \lambda \geq 0$,

$$|\varphi_0(x, y) - \varphi_0(x', y)| \leq \kappa_1(\rho, \lambda)\|x - x'\| \quad \forall x, x' \in \mathbb{R}_+, \ y \in B_\lambda,$n

$$|\varphi_0(x, y) - \varphi_0(x, y')| \leq \kappa_2(\rho, \lambda)d_Y(y, y') \quad \forall x \in \mathbb{R}_+, \ y, y' \in B_\lambda,$n

where, for some centroid $\bar{y} \in Y$, $B_{\lambda} := B(\bar{y}, \lambda)$. Again, as in the case of the centroid $\bar{x} \in X$, $\bar{y}$ serves as a centroid of $Y$ and can be selected arbitrary. However, it is beneficial in practice to make the choice in view of the “scale” and “location” of the optimal maximization problem. We say that a set-valued mapping $D_0 : \mathbb{R}^n \Rightarrow Y$ is uniformly $\hat{d}_{\lambda}$-Lipschitz continuous with modulus $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ if, for all $\rho \geq 0$ and $\lambda \geq 0$,

$$\hat{d}_{\lambda}(D_0(x), D_0(x')) \leq \gamma(\rho)||x - x'|| \quad \forall x, x' \in \rho B.$n

This is a strong property, but it holds with $\gamma(\rho) = 0$ if $D_0$ is independent of $x$.

Theorem 5.6 (bound under Lipschitz continuity). Suppose $\varphi, \varphi' \in \text{bfcns(}\mathbb{R}^n, Y)$ have domains given by closed $C, C'$ and closed-valued mappings $D, D'$ and, in terms of marginally Lipschitz continuous bifunctions $\varphi_0, \varphi'_0 : \mathbb{R}^n \times Y \to \mathbb{R}$, with common moduli $\kappa_1, \kappa_2 : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$, are given by

$$\varphi = \varphi_0 \text{ on } \text{dom } \varphi \quad \text{and} \quad \varphi' = \varphi'_0 \text{ on } \text{dom } \varphi'.$n

Let the set-valued mappings $D_0, D'_0 : \mathbb{R}^n \Rightarrow Y$ be closed- and nonempty-valued and uniformly $\hat{d}_{\lambda}$-Lipschitz continuous with common modulus $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$. Suppose

$$D = D_0 \text{ on } C \quad \text{and} \quad D' = D'_0 \text{ on } C'.$n

For a fixed $\rho \geq 0$, suppose that for some $\rho^* \geq \rho + \hat{d}_\rho(C, C')$, $\lambda^* \geq 0$, and $\sigma < \infty$, we have that

(i) $d_{\lambda, \rho}(D_0(x), D'_0(x)) \leq \sigma$ for all $x \in \rho B$,

(ii) $\argmax_{D_0(x)} \varphi_0(x, \cdot) \cap B_{\lambda} \neq \emptyset$ and $\argmax_{D'_0(x)} \varphi'_0(x, \cdot) \cap B_{\lambda} \neq \emptyset$ for $x \in \rho^* B$,

(iii) $\sup_{D_0(x)} \varphi_0(x, \cdot), \sup_{D'_0(x)} \varphi'_0(x, \cdot) \in [-\lambda^*, \lambda^*]$ if $x \in \rho^* B$ and finite otherwise.

Then, the sup-projections $h, h'$ of $\varphi, \varphi'$, respectively, exist and satisfy

$$\hat{d}_\rho^\prime(h, h') \leq \sup_{x \in \rho B, \ y \in B_{\lambda'} \ (\ i)} |\varphi_0(x, y) - \varphi'_0(x, y)|$$

$$+ \kappa_1 \hat{d}_\rho(C, C') + \kappa_2 \sup_{x \in \rho B} \hat{d}_{\lambda'}(D_0(x), D'_0(x)).$$
Thus, by Proposition 5.5, for $x, x'$ an application of Proposition 5.5. For any $\rho$, sequence of repeated applications of Propositions 5.5 and 5.3. Fix $D$ and $\lambda$. Utilizing the bound (4) on the right-hand side gives that

$$\sup_{y \in D_0(x)} \varphi_0(x, y) \leq \kappa_2(\rho^*, \lambda^*) \max \{1, \kappa_1(\rho^*, \lambda^*) + \max \{1, \kappa_2(\rho^*, \lambda^* + \gamma(\rho^*)\rho^* + \varepsilon)\} \}$$

whence $\bar{\kappa}_1 = \max \{1, \kappa_1(\rho^*, \lambda^*) + \max \{1, \kappa_2(\rho^*, \lambda^* + \gamma(\rho^*)\rho^* + \varepsilon)\} \}$. $\bar{\kappa}_2 = \max \{1, \kappa_2(\rho, \lambda^* + \sigma + \varepsilon)\}$. $\bar{\kappa}_2 = \max \{1, \kappa_2(\rho, \lambda^* + \sigma + \varepsilon)\}$ and $\varepsilon > 0$. When $Y$ is proper, then $\varepsilon = 0$ is permissible.

Proof. For any $\rho$ and $\lambda$, let $\kappa_2(\rho, \lambda) = \max \{1, \kappa_2(\rho, \lambda)\}$. The result is a consequence of repeated applications of Propositions 5.5 and 5.3. Fix $\rho \geq 0$, and let $\rho^* \geq \rho + \overline{d}_\rho(C, C')$ be such that the assumptions of the theorem hold. We start with an application of Proposition 5.5. For any $x \in \rho^* B$, $D_0(x)$ is closed and nonempty, and $\varphi_0(x, \cdot) : Y \rightarrow \mathbb{R}$ satisfies

$$|\varphi_0(x, y) - \varphi_0(x, y')| \leq \kappa_2(\rho^*, \lambda) d_Y(y, y')$$

for any $\lambda \geq 0$ and $y, y' \in B^\lambda$.

Thus, by Proposition 5.5, for $x, x' \in \rho^* B$ and $\lambda \geq 0$,

$$\eta_\lambda(-\varphi_0(x, \cdot) : D_0(x) \rightarrow \mathbb{R}, -\varphi_0(x', \cdot) : D_0(x') \rightarrow \mathbb{R})$$

when $\lambda' \geq \lambda + \overline{d}_\lambda(D_0(x), D_0(x')) + \varepsilon$. For $x, x' \in \rho^* B$ and $\lambda \geq 0$, the Lipschitz continuity of $D_0$ implies that $\overline{d}_\lambda(D_0(x), D_0(x')) \leq \gamma(\rho^*)\lambda^* \leq \gamma(\rho^*)\rho^*$. These facts and the Lipschitz continuity of $\varphi_0(\cdot, y)$ establish that for $x, x' \in \rho^* B$ and $\lambda \geq 0$,

$$\eta_\lambda(-\varphi_0(x, \cdot) : D_0(x) \rightarrow \mathbb{R}, -\varphi_0(x', \cdot) : D_0(x') \rightarrow \mathbb{R})$$

(4) $$\leq \left[\kappa_1(\rho^*, \lambda) + \kappa_2(\rho^*, \lambda + \gamma(\rho^*)\rho^* + \varepsilon)\gamma(\rho^*)\right]\|x - x'\|_2.$$

A similar argument gives the same result for $\varphi_0^\prime$ and $D_0^\prime$.

We next apply Proposition 5.3, reoriented towards maximization instead of minimization, and find that for $x, x' \in \rho^* B$ and the sufficiently large $\lambda^*$,

$$\sup_{y \in D_0(x)} \varphi_0(x, y) - \sup_{y \in D_0(x')} \varphi_0(x', y)$$

$$\leq \eta_\lambda(-\varphi_0(x, \cdot) : D_0(x) \rightarrow \mathbb{R}, -\varphi_0(x', \cdot) : D_0(x') \rightarrow \mathbb{R}).$$

Utilizing the bound (4) on the right-hand side gives that

$$\left|\sup_{y \in D_0(x)} \varphi_0(x, y) - \sup_{y \in D_0(x')} \varphi_0(x', y)\right|$$

$$\leq \left[\kappa_1(\rho^*, \lambda^*) + \kappa_2(\rho^*, \lambda^* + \gamma(\rho^*)\rho^* + \varepsilon)\gamma(\rho^*)\right]\|x - x'\|_2 \quad \forall x, x' \in \rho^* B.$$

A similar argument establishes the same Lipschitz property for $\sup_{y \in D_0^\prime(\cdot, y)} \varphi_0^\prime(\cdot, y)$ on $\rho^* B$.

We next apply Proposition 5.5 for a second time. Under the stated assumptions, $\sup_{y \in D_0(\cdot, y)} \varphi_0(\cdot, y)$ and $\sup_{y \in D_0^\prime(\cdot, y)} \varphi_0^\prime(\cdot, y)$ are finite-valued on $\mathbb{R}^\ast$. Moreover, they satisfy the previously established Lipschitz property on $\rho^* B$. Thus, by Proposition 5.5,

$$\overline{d}_\rho(h, h') \leq \max_{x \in \rho^* B} \left|\sup_{y \in D_0(x)} \varphi_0(x, y) - \sup_{y \in D_0^\prime(x)} \varphi_0^\prime(x, y)\right|$$

(5) $$\overline{d}_\rho(h, h') \leq \max_{x \in \rho^* B} \left|\sup_{y \in D_0(x)} \varphi_0(x, y) - \sup_{y \in D_0^\prime(x)} \varphi_0^\prime(x, y)\right|$$

$$+ \max \left\{1, \kappa_1(\rho^*, \lambda^*) + \kappa_2(\rho^*, \lambda^* + \gamma(\rho^*)\rho^* + \varepsilon)\gamma(\rho^*)\right\}\overline{d}_\rho(C, C').$$
It remains to bound the first term after the inequality. An intermediate step is a third application of Proposition 5.5. For \( x \in \rho \mathcal{B} \),

\[
(6) \quad \eta_\lambda \left( - \varphi_0(x, \cdot) : D_0(x) \to \mathbb{R}, -\varphi_0^* (x, \cdot) : D_0^* (x) \to \mathbb{R} \right) \\
\leq \sup_{y \in \mathcal{B}_\lambda} \left| \varphi_0 (x, y) - \varphi_0^* (x, y) \right| \leq \kappa_2 (\rho, \lambda + \sigma + \varepsilon) \hat{d}_\lambda (D_0(x), D_0^*(x)).
\]

Another application of Proposition 5.3 yields the bound

\[
\left| \sup_{y \in D_0(x)} \varphi_0 (x, y) - \sup_{y \in D_0^*(x)} \varphi_0^* (x, y) \right| \\ \leq \eta_\lambda \left( - \varphi_0(x, \cdot) : D_0(x) \to \mathbb{R}, -\varphi_0^* (x, \cdot) : D_0^* (x) \to \mathbb{R} \right) \quad \text{for } x \in \rho \mathcal{B}.
\]

Combining this result with (6), we obtain that

\[
\max_{x \in \rho \mathcal{B}} \sup_{y \in D_0(x)} \varphi_0 (x, y) - \sup_{y \in D_0^*(x)} \varphi_0^* (x, y) \\ \leq \sup_{x \in \rho \mathcal{B}, y \in \mathcal{B}_\lambda} \left| \varphi_0 (x, y) - \varphi_0^* (x, y) \right| + \kappa_2 (\rho, \lambda + \sigma + \varepsilon) \sup_{x \in \rho \mathcal{B}} \hat{d}_\lambda (D_0(x), D_0^*(x)).
\]

In view of this bound and (5), the conclusion follows.

When \( D, D^\nu \) are independent of \( x \), the Lipschitz constants are uniform, and other assumptions are made, the result simplifies considerably.

**Corollary 5.7.** Suppose that \( \varphi, \varphi^\nu \in \text{bfcns}(\mathbb{R}^m, Y) \) have domains given by closed \( C, C^\nu \subset \mathbb{R}^m \) and compact \( D, D^\nu \subset Y \) and, in terms of marginally Lipschitz continuous bifunctions \( \varphi_0, \varphi_0^\nu : \mathbb{R}^m \times Y \to \mathbb{R} \), with common moduli \( \kappa_1, \kappa_2 \) independent of \( \rho, \lambda \), are given by

\[
\varphi = \varphi_0 \text{ on } C \times D \quad \text{and} \quad \varphi^\nu = \varphi_0^\nu \text{ on } C^\nu \times D^\nu.
\]

For \( \rho \geq 0 \), suppose that, for some \( \rho^\star \geq \rho + \hat{d}_\rho (C, C^\nu) \) and \( \lambda^\star \geq 0 \), \( \arg\max_D \varphi_0 (x, \cdot) \cap \mathcal{B}_\lambda \neq \emptyset \), \( \arg\max_{D^\nu} \varphi_0^\nu (x, \cdot) \cap \mathcal{B}_\lambda \neq \emptyset \), and \( \max_D \varphi_0 (x, \cdot), \max_{D^\nu} \varphi_0^\nu (x, \cdot) \in [-\lambda^\star, \lambda^\star] \) for all \( x \in \rho \mathcal{B} \) and finite elsewhere. Then, the sup-projections \( h, h^\nu \) of \( \varphi, \varphi^\nu \), respectively, exist and

\[
\hat{d}_\rho (h, h^\nu) \leq \sup_{x \in \rho \mathcal{B}, y \in \mathcal{B}_\lambda} \left| \varphi_0 (x, y) - \varphi_0^\nu (x, y) \right| \\
+ \max \{1, \kappa_1\} \hat{d}_\rho (C, C^\nu) + \max \{1, \kappa_2\} \hat{d}_\lambda (D, D^\nu).
\]

**5.3. Examples of solution estimates.** We end the paper with two illustrations of estimates of the distance between minsup-values. We start with a result of independent interest.

**Proposition 5.8 (distance under stochastic dominance constraints).** Suppose that \( F_0, G_0, F_0^\nu, G_0^\nu \in \text{cd-fcns}(\mathbb{R}^m) \) are reference distribution functions and

\[
S = \{ F \in \text{cd-fcns}(\mathbb{R}^m) : F_0 (\xi) \leq F (\xi) \leq G_0 (\xi) \ \forall \xi \in \mathbb{R}^m \}, \quad \text{and} \quad S^\nu = \{ F \in \text{cd-fcns}(\mathbb{R}^m) : F_0^\nu (\xi) \leq F (\xi) \leq G_0^\nu (\xi) \ \forall \xi \in \mathbb{R}^m \}
\]

are nonempty subsets of \( \text{use-fcns}(\mathbb{R}^m; [0, 1]), \hat{d}_b \). Then, for any \( \rho \geq 0 \) and \( \lambda \geq 0 \),

\[
\hat{d}_\rho (S, S^\nu) \leq (1 - e^{-\lambda}) \max \{ \hat{d}_\rho (F_0, F_0^\nu), \hat{d}_\lambda (G_0, G_0^\nu) \} + e^{-\lambda} (\lambda + 1).
\]
Moreover,

\[ \hat{d}_\rho(S, S') \leq d^h(F_0, F'_0) \text{ if } G_0 = G'_0 \quad \text{and} \quad \hat{d}_\rho(S, S') \leq d^h(G_0, G'_0) \text{ if } F_0 = F'_0. \]

Proof. Let \( F \in S \), and construct \( F'(\xi) = \max\{F_0(\xi), \min[F(\xi), G_0(\xi)]\} \) for \( \xi \in \mathbb{R}^m \), which is a distribution function in \( S' \). Since \( \text{hypo } F_0 \subset \text{hypo } F \subset \text{hypo } G_0 \) and similarly for \( F'_0, F', G'_0 \), we have for every \( \lambda \geq 0 \), \( d^h_F(F, F') \leq \max\{d^h\rho(F_0, F'_0), d^h\rho(G_0, G'_0)\} \). From Proposition 5.1 and its proof one can then deduce that

\[ d^h(F, F') \leq (1 - e^{-\lambda}) \max\{d^h\rho(F_0, F'_0), d^h\rho(G_0, G'_0)\} + e^{-\lambda}(\lambda + 1). \]

Repeating this argument with the roles of \( F \) and \( F' \) reversed leads to the first conclusion. The second and third conclusions are immediate, as then the max in the second-to-last expression is eliminated.

The previous result, in conjunction with Theorem 5.6 and Corollary 5.7, provides a central tool for bounding errors in minsup-value in settings such as those in Example 10.

Example 13: Superquantile-risk minimization (cont.). Since \( D(x) \) and \( D'(x) \) differ only by \( \alpha(x) \) being replaced by \( \alpha'(x) \), it is clear that for all \( q \in D(x) \) there exists \( q' \in D'(x) \) with

\[ \|q - q'\|_\infty \leq \left| \frac{1}{1 - \alpha(x)} - \frac{1}{1 - \alpha'(x)} \right| = \frac{|\alpha(x) - \alpha'(x)|}{(1 - \alpha(x))(1 - \alpha'(x))}. \]

Since this holds also with the roles of \( q \) and \( q' \) reversed, we obtain that for all \( \rho \geq 0 \)

\[ \hat{d}_\rho(D(x), D'(x)) \leq \frac{|\alpha(x) - \alpha'(x)|}{(1 - \alpha(x))(1 - \alpha'(x))}. \]

For \( q, q' \in \mathcal{L}_\infty(\mathbb{R}^m), |\varphi(x, q) - \varphi(x, q')| \leq \int |\psi(x, \xi)|dP_\rho(\xi)||q - q'||_\infty, \) which establishes the required Lipschitz constant in Theorem 5.6 or Corollary 5.7. Thus, if there is no approximation of the cost function and \( C = C' \), then Theorem 5.6 and Proposition 5.3 lead, for sufficiently large \( \rho \), to

\[ \minsup \varphi - \minsup \varphi' \leq \max \left\{ 1, \sup_{x \in \mathbb{R}^n} \int |\psi(x, \xi)|dP_\rho(\xi) \right\} \sup_{x \in \mathbb{R}^n} \frac{|\alpha(x) - \alpha'(x)|}{(1 - \alpha(x))(1 - \alpha'(x))}. \]

This expression provides a bound on the price of robustness, which increases at a linear rate as, say, \( \alpha \) moves away from a nominal \( \alpha = 0 \).

Appendix. Proof of Theorem 3.5. Since \( d^h(F, G) = d^e(-F, -G) \), we follow an argument that is similar to that of Proposition 5.1. Utilizing the facts that \( \text{dist}(0, \text{epi}(-F)) = \text{dist}(0, \text{epi}(-G)) = 0 \) and \( F \) and \( G \) are both bounded between 0 and 1, a line of arguments similar to those in the proof of Proposition 5.1 leads to

\[ e^{-\rho}d^e_\rho(-F, -G) \leq d^e(-F, -G) \leq e^{-\rho} + (1 - e^{-\rho})d^2_\rho(-F, -G). \]

Thus, we need only to construct a lower bound on \( d^e_\rho(-F, -G) \) and an upper bound on \( d^2_\rho(-F, -G) \). In both cases, we utilize Proposition 5.2.
First, we consider the lower bound. Since \( F, G \leq 1 \), a \( \rho \geq 1 \) simplifies the alternative expression for \( \hat{d}^e_\rho(-F, -G) \) in Proposition 5.2 to

\[
\hat{d}^e_\rho(-F, -G) = \inf \left\{ \eta \geq 0 : \min_{B(\xi, \eta)} -G \leq -F(\xi) + \eta, \min_{B(\xi, \eta)} -F \leq -G(\xi) + \eta \forall \xi \in \rho \mathbb{B} \right\}.
\]

Replacing the minimization over a ball by minimization over the smallest hypercube containing the ball, we obtain a relaxation of the infimum problem over \( \eta \). Due to the monotonicity of \( F \) and \( G \), the minimization over the hypercube is attained at a particular vertex. Hence, for \( \rho \geq 1 \),

\[
\hat{d}^e_\rho(-F, -G) \geq \inf \left\{ \eta \geq 0 : G(\xi + \eta 1) + \eta \geq F(\xi), F(\xi + \eta 1) + \eta \geq G(\xi) \forall \xi \in \rho \mathbb{B} \right\}.
\]

Second, we consider an upper bound on \( \hat{d}^e_{2\rho}(-F, -G) \). Similar to the case for the lower bound, because \( F, G \leq 1 \), a \( \rho \geq 1/2 \) simplifies the expression for \( \hat{d}^e_{2\rho}(-F, -G) \) to

\[
\hat{d}^e_{2\rho}(-F, -G) = \inf \left\{ \eta \geq 0 : \min_{B(\xi, \eta)} -G \leq -F(\xi) + \eta, \min_{B(\xi, \eta)} -F \leq -G(\xi) + \eta \forall \xi \in 2\rho \mathbb{B} \right\}.
\]

Replacing the minimization over a ball by minimization over the largest hypercube contained in the ball, we obtain a restriction of the infimum problem over \( \eta \). Due to the monotonicity of \( F \) and \( G \), the minimization over the hypercube is attained at a particular vertex. Hence, for \( \rho \geq 1/2 \),

\[
\hat{d}^e_{2\rho}(-F, -G) \leq \inf \left\{ \eta \geq 0 : G(\xi + \eta 1/\sqrt{m}) + \eta \geq F(\xi), F(\xi + \eta 1/\sqrt{m}) + \eta \geq G(\xi) \forall \xi \in 2\rho \mathbb{B} \right\}.
\]

Denoting the lower bounding and upper bounding expressions by \( \eta(\rho) \) and \( \bar{\eta}(\rho) \), respectively, yields the first two inequalities. Letting \( \rho \to \infty \) in the upper bound, we obtain the last inequality. \( \square \)

REFERENCES


