MODEL-FREE BOUNDS FOR MULTI-ASSET OPTIONS USING OPTION-IMPLIED INFORMATION AND THEIR EXACT COMPUTATION

ARIEL NEUFELD, ANTONIS PAPAPANTOLEON, AND QIKUN XIANG

ABSTRACT. We consider derivatives written on multiple underlyings in a one-period financial market, and we are interested in the computation of model-free upper and lower bounds for their arbitrage-free prices. We work in a completely realistic setting, in that we only assume the knowledge of traded prices for other single- and multi-asset derivatives, and even allow for the presence of bid-ask spread in these prices. We provide a fundamental theorem of asset pricing for this market model, as well as a superhedging duality result, that allows to transform the abstract maximization problem over probability measures into a more tractable minimization problem over vectors, subject to certain constraints. Then, we recast this problem into a linear semi-infinite optimization problem, and provide two algorithms for its solution. These algorithms provide upper and lower bounds for the prices that are ε-optimal, as well as a characterization of the optimal pricing measures. Moreover, these algorithms are efficient and allow the computation of bounds in high-dimensional scenarios (*e.g.* when $d = 60$). Numerical experiments using synthetic data showcase the efficiency of these algorithms, while they also allow to understand the reduction of model-risk by including additional information, in the form of known derivative prices.

1. INTRODUCTION

The classical paradigm in finance and theoretical economics assumes the existence of a model that provides an accurate description of the evolution of asset prices, and all subsequent computations about hedging strategies, exotic derivatives, risk measures, and so forth, are based on this model. However, academics, practitioners, and regulators have realized that all models provide only a partially accurate description of this reality, thus, either methods need to be developed in order to aggregate the results of many models, or approaches have to be devised that allow for computations in the absence of a specific model. The first approach led to the introduction of robust methods in asset pricing and no-arbitrage theory, see *e.g.* Bayraktar, Huang, and Zhou [7], Beissner [8], Beissner and Riedel [9, 10], Bouchard and Nutz [15], Bouchard, Deng, and Tan [16], Dana and Riedel [24], Epstein and Ji [35], Neufeld and Nutz [50], Rigotti and Shannon [58] and Yan, Cheng, Natarajan, and Teo [63], while the second one led to model-free methods in asset pricing and no-arbitrage theory, see *e.g.* Acciaio, Beiglböck, Penkner, and Schachermayer [1], Bartl, Kupper, Prömel, and Tangpi [5], Bartl, Kupper, and Neufeld [6], Burzoni, Frittelli, and Maggis [17], Burzoni, Riedel, and Soner [18], Burzoni, Frittelli, Hou, Maggis, and Obłój [19], Cheridito, Kupper, and Tangpi [21], Dolinsky and Neufeld [30], Dolinsky and Soner [31], Galichon, Henry-Labordere, and Touzi [36], Hobson [42], Hu, Li, and Mehrotra [43], Lütkebohmert and Sester [48], Possamaï, Royer, and Touzi [53] and Riedel [57].

In this work, we consider derivatives written on multiple underlyings in a one-period financial market, and we are interested in the computation of upper and lower bounds for their arbitrage-free prices. We work in a completely realistic setting, in that we only assume the knowledge of traded prices for other single- and multi-asset derivatives, and even allow for the presence of bid-ask

2010 *Mathematics Subject Classification.* 91G20, 91G60, 90C05.

Key words and phrases. Model-free bounds, option-implied information, multi-asset options, bid-ask spread, cutting plane method, no-arbitrage gap.

We thank Daniel Bartl, Julien Guyon, and Steven Vanduffel for fruitful comments and discussions that initiated this project. We are also grateful to Stephan Eckstein and Michael Kupper for fruitful discussions during the work on these topics. AN gratefully acknowledges the financial support by his NAP Grant. AP gratefully acknowledges the financial support from the Hellenic Foundation for Research and Innovation Grant No. HFRI-FM17-2152.

spread in these prices. In other words, we work in a model-free setting in the presence of option-implied information, and make no assumption about the probabilistic evolution of asset prices (*i.e.* their marginal distributions) or their dependence structure.

The computation of bounds for the prices of multi-asset options, most often basket options, is a classical problem in the mathematical finance literature and has connections with several other branches of mathematics, such as probability theory, optimal transport, operations research, and optimization. In the most classical setting, one assumes that the marginal distributions are known and the joint law is unknown; this framework is known as dependence uncertainty. In this framework, several authors have derived bounds for multi-asset options using tools from probability theory, such as copulas and Fréchet–Hoeffding bounds, see *e.g.* Chen, Deelstra, Dhaene, and Vanmaele [20], Dhaene, Denuit, Goovaerts, Kaas, and Vyncke [28, 29] and Hobson, Laurence, and Wang [40, 41]. These bounds turned out to be very wide for practical applications, hence recently there was an interest in methods that allow for the inclusion of additional information on the dependence structure, in order to reduce this gap. This led to the creation of improved Fréchet–Hoeffding bounds and the pricing of multi-asset options in the presence of additional information on the dependence structure, see *e.g.* Tankov [61], Lux and Papapantoleon [49] and Puccetti, Rüschendorf, and Manko [54].

The setting of dependence uncertainty is intimately linked with optimal transport theory, and its tools have also been used in order to derive bounds for multi-asset option prices, see *e.g.* Bartl, Kupper, Lux, Papapantoleon, and Eckstein [4] for a formulation in the presence of additional information on the joint distribution. More recently, Aquino and Bernard [2], Eckstein and Kupper [32] and Eckstein, Guo, Lim, and Obłój [34] have translated the model-free superhedging problem into an optimization problem over classes of functions, by extending results in optimal transport, and used neural networks and the stochastic gradient descent algorithm for the computation of the bounds. We refer the reader to Section 3.3 for more information and a comparison with our approach.

Ideas from operations research and optimization have also been applied for the computation of model-free bounds in settings that are closer to ours, and do not necessarily assume knowledge of the marginal distributions (or, equivalently, knowledge of call option prices for a continuum of strikes). d'Aspremont and El Ghaoui [25] consider a framework where the prices of forwards and single-asset call options are known, and compute upper and lower bounds on basket options prices using linear programming. In the more general case where the prices of other basket options are also known, they derive a relaxation to the problem which can be solved using linear programming. This seminal work was later extended by various authors. Peña, Vera, and Zuluaga [52] improve the results of [25] when computing the lower bounds on basket options prices in two special cases: (i) when the number of assets is limited to two and prices of basket options are known; and (ii) when the prices of only a forward and a single-asset call option per asset are known. Peña, Saynac, Vera, and Zuluaga [51] study the problem of computing the upper and lower bounds on basket and spread option prices when the asset prices are non-negative and upper-bounded, and prices of other basket and spread option prices are known. Their approach involves solving a large linear programming problem via the Dantzig–Wolfe decomposition in which the corresponding subproblem is solved using mixed-integer programming. Compared to [25, 51, 52], the numerical methods we develop in Section 3 apply to settings that are much more general, where the derivative being priced and the traded derivatives with known prices can be any continuous piece-wise affine function (including but not limited to vanilla, basket, spread, and rainbow options, as well as any linear combination of these options). Moreover, as we demonstrate in Section 4, these methods are able to efficiently compute the price bounds in high-dimensional scenarios, *e.g.* when 60 assets are considered. This is considerably higher compared to existing studies. Daum and Werner [26] develop a discretization-based algorithm for solving linear semi-infinite programming problems that returns a feasible solution, and apply the algorithm to compute the upper and lower bounds on basket or spread options prices when single-asset call, put, and exotic options prices are known. The algorithm we introduce in Section 3.1 takes a similar approach, but is able to solve the problem when the prices of multi-asset options are known. Kahalé [44] uses a central cutting plane algorithm to compute the super- and sub-replicating prices of financial derivatives using hedging portfolios that consist of other financial derivatives in the multi-period discrete-time setting. The algorithm only works under the assumption that the underlying state space (*i.e.* the space of asset prices) is finite. When the state space is infinite, it is discretized before applying the central cutting plane algorithm, and the discretization error is analyzed. However, the approach of discretizing the state space has limited applicability to the multi-dimensional settings (*i.e.* with multiple underlying assets) due to the curse of dimensionality. The algorithm we develop in Section 3.2 is also based on a central cutting plane algorithm, but it allows us to efficiently compute model-free price bounds in high-dimensional state spaces for financial derivatives that depend on multiple assets.

Our contributions are three-fold: Firstly, we provide a fundamental theorem of asset pricing for the market model described above, as well as a superhedging duality, that allows to transform the abstract maximization problem over probability measures into a more tractable problem over vectors, subject to certain constraints. Secondly, we recast this problem into a linear semi-infinite optimization problem, and provide two algorithms for its solution. These algorithms provide upper and lower bounds for the prices of multi-asset derivatives that are ε-optimal, as well as a characterization of the optimal pricing measures. Moreover, these algorithms are efficient and allow the computation of bounds in high-dimensional scenarios (*e.g.* when $d = 60$). Thirdly, we perform numerical experiments using synthetic data that showcase the efficiency of these algorithms. These experiments allow us to understand the reduction of the no-arbitrage gap, *i.e.* the difference between the upper and lower no-arbitrage bounds, by including additional information in the form of known derivative prices. The no-arbitrage gap directly reflects the model-risk associated to a particular derivative and the information available in the market. The numerical experiments show a decrease of the model-risk by the inclusion of additional information, although this decrease is not uniform and depends on the form of information and the specific structure of the payoff functions.

This paper is organized as follows: In Section 2, we present the modeling framework, state the no-arbitrage theorem and the superhedging duality, and discuss a setting that is relevant for practical applications. In Section 3 we present the algorithms that have been developed for the computation of model-free bounds, comment on the choice of the different parameters of the algorithms, and state the theorems that show the validity of these algorithms. In Section 4, we discuss various numerical experiments, both in low ($d = 5$) and in high ($d = 60$) dimensions, that show the efficiency of the algorithms and the reduction of model-risk by the inclusion of additional information in the form of known derivative prices. Finally, the Appendices contain the proofs of the results in Sections 2 and 3.

2. Duality in the presence of option-implied information

In this section, we introduce a general framework for a model-free, one-period, financial market where multiple assets and several derivatives written on these assets are traded simultaneously. Model-free means that we will not make any assumption about the probabilistic model that governs the evolution of asset prices. Instead, we will utilize information available in the financial market and implied by single- and multi-asset option prices. We will provide both a fundamental theorem and a superhedging duality in this setting, where our results and proofs are inspired by Bouchard and Nutz [15]. Moreover, we will describe concrete examples of this framework that are of practical interest.

Throughout this work, all vectors are column vectors unless otherwise stated. We denote vectors and vector-valued functions by boldface symbols. Let $\boldsymbol{x}$ be a vector, then $[\boldsymbol{x}]_j$ denotes the j-th component of $\boldsymbol{x}$. For simplicity, we also use x_j to denote $[\boldsymbol{x}]_j$ when there is no ambiguity. For $i < j$, $[\boldsymbol{x}]_{i:j}$ denotes the vector formed with the i-th through the j-th entries of $\boldsymbol{x}$, *i.e.* $[\boldsymbol{x}]_{i:j} := (x_i, x_{i+1}, \ldots, x_j)^\mathsf{T}$. Let $\|\boldsymbol{x}\|$ denote the Euclidean norm of $\boldsymbol{x}$. Let $\boldsymbol{x}$ and $\boldsymbol{x}'$ be two vectors, then $\langle \boldsymbol{x}, \boldsymbol{x}' \rangle$ denotes their inner-product. We denote by $\boldsymbol{e}_i$ the i-th standard basis vector of $\mathbb{R}^d$, by $\mathbf{0}$ the vector with all entries equal to zero, *i.e.* $\mathbf{0} = (0, \ldots, 0)^\mathsf{T}$, and by $\mathbf{1}$ the vector with all entries equal to one, *i.e.* $\mathbf{1} = (1, \ldots, 1)^\mathsf{T}$. We call a subset of Euclidean space a polyhedron or a polyhedral set if it is the intersection of finitely many closed half-spaces. We call a subset of Euclidean space a polytope if it is a bounded polyhedron. A convex subset C' of a convex set $C \subset \mathbb{R}^d$ is called a face of C if

$$\boldsymbol{x} \in C' \text{ and } \boldsymbol{x} = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \implies \boldsymbol{x}_1 \in C', \boldsymbol{x}_2 \in C' \quad \text{for } 0 < \lambda < 1, \boldsymbol{x}_1 \in C, \boldsymbol{x}_2 \in C.$$

Let $A \subset \mathbb{R}^d$, then $\mathrm{cone}(A)$ denotes the conic hull of the set A. We refer the reader to Rockafellar [60] for further standard notions and notations from convex analysis.

Let Ω be a Polish space equipped with its Borel σ-algebra denoted by $\mathcal{B}(\Omega)$. Let $\mathcal{P}(\Omega)$ denote the set of Borel probability measures on Ω. Let $g_j : \Omega \mapsto \mathbb{R}$ be Borel measurable for $j = 1, \ldots, m$. Let $\boldsymbol{g} : \Omega \mapsto \mathbb{R}^m$ denote the vector-valued Borel measurable function where the j-th component corresponds to g_j. Let $\underline{\pi}_j, \overline{\pi}_j \in \mathbb{R}$ be such that $\underline{\pi}_j \leq \overline{\pi}_j$ for $j = 1, \ldots, m$. Let Δ denote the following polytope,

$$\Delta := [\underline{\pi}_1, \overline{\pi}_1] \times \cdots \times [\underline{\pi}_m, \overline{\pi}_m] \subset \mathbb{R}^m. \tag{2.1}$$

Let $\boldsymbol{y} = (y_1, \ldots, y_m)^\mathsf{T} \in \mathbb{R}^m$ and define $\pi : \mathbb{R}^m \mapsto \mathbb{R}$ as

$$\pi(\boldsymbol{y}) := \sum_{j=1}^{m} y_j^+ \overline{\pi}_j - y_j^- \underline{\pi}_j, \tag{2.2}$$

where $y_j^+ := \max\{y_j, 0\}, y_j^- := \max\{-y_j, 0\}$. One verifies from definitions (2.1) and (2.2) that

$$\pi(\boldsymbol{y}) = \max_{\boldsymbol{q} \in \Delta} \langle \boldsymbol{y}, \boldsymbol{q} \rangle, \tag{2.3}$$

thus $\pi(\cdot)$ is sublinear. Let $\langle \boldsymbol{y}, \boldsymbol{g} \rangle$ denote the function $\sum_{j=1}^m y_j g_j : \Omega \mapsto \mathbb{R}$.

We make the following *no-arbitrage* assumption.

Assumption 2.1 (No-arbitrage)**.** The following implication holds for any $\boldsymbol{y} \in \mathbb{R}^m$:

$$\langle \boldsymbol{y}, \boldsymbol{g} \rangle - \pi(\boldsymbol{y}) \geq 0 \quad \Longrightarrow \quad \langle \boldsymbol{y}, \boldsymbol{g} \rangle - \pi(\boldsymbol{y}) = 0,$$

where the inequality and the equality are both understood as point-wise.

Remark 2.2. Assumption 2.1 is a special case of the no-arbitrage assumption introduced in Definition 1.1 of Bouchard and Nutz [15], where the set of possible models for the market is $\mathcal{P}(\Omega)$, *i.e.* all Borel probability measures, and a single time step is considered. The difference between Assumption 2.1 and the no-arbitrage assumption in [15] is that the price of a financial derivative in the present work is not a singleton but can lie anywhere between the corresponding bid and ask prices $\underline{\pi}, \overline{\pi}$. Note that there are other notions of no-arbitrage that are weaker than Assumption 2.1, for example the "no uniform strong arbitrage" assumption in Definition 2.1 of Bartl et al. [4].

Let $f : \Omega \mapsto \mathbb{R}$ be a Borel measurable function, and define the functional $\phi(f)$ as follows:

$$\phi(f) := \inf \big\{ c + \pi(\boldsymbol{y}) : c \in \mathbb{R}, \boldsymbol{y} \in \mathbb{R}^m, c + \langle \boldsymbol{y}, \boldsymbol{g} \rangle \geq f \big\}. \tag{2.4}$$

Let $\mathcal{Q}$ be defined as follows:

$$\mathcal{Q} := \left\{ \mu \in \mathcal{P}(\Omega) : \underline{\pi}_j \leq \int_\Omega g_j \,\mathrm{d}\mu \leq \overline{\pi}_j, \text{ for } j = 1, \ldots, m \right\}. \tag{2.5}$$

Moreover, define the sets $\widetilde{\mathcal{P}} \subset \mathcal{P}(\Omega)$ and $\Gamma \subset \mathbb{R}^m$ as follows:

$$\widetilde{\mathcal{P}} := \left\{ \mu \in \mathcal{P}(\Omega) : \int_\Omega |f| + \textstyle\sum_{j=1}^m |g_j| \,\mathrm{d}\mu < \infty \right\}, \tag{2.6}$$

$$\Gamma := \left\{ \int_\Omega \boldsymbol{g} \,\mathrm{d}\mu : \mu \in \widetilde{\mathcal{P}} \right\}, \tag{2.7}$$

where $\int_\Omega \boldsymbol{g} \,\mathrm{d}\mu$ denotes $\big(\int_\Omega g_1 \,\mathrm{d}\mu, \ldots, \int_\Omega g_m \,\mathrm{d}\mu \big)^\mathsf{T} \in \mathbb{R}^m$.

The main results of this section are the following fundamental theorem and super-hedging duality, similar to Bouchard and Nutz [15].

Theorem 2.3 (Fundamental Theorem)**.** *The following are equivalent:*

(i) Assumption 2.1 holds.
(ii) For all $\nu \in \mathcal{P}(\Omega)$, there exists $\mu \in \mathcal{Q}$ such that $\nu \ll \mu$.

Proof. See Appendix A. □

Theorem 2.4 (Superhedging duality)**.** *Under Assumption 2.1, the following statements hold:*

(i) $\phi(f) > -\infty$.

(ii) There exists $\boldsymbol{y} \in \mathbb{R}^m$ *such that* $\phi(f) + \langle \boldsymbol{y}, \boldsymbol{g} \rangle - \pi(\boldsymbol{y}) \geq f$. *Hence, the infimum in (2.4) is attained when* $\phi(f) < \infty$.

(iii) We have the following superhedging duality result:

$$\phi(f) = \sup_{\mu \in \mathcal{Q}} \int_{\Omega} f \, \mathrm{d}\mu. \tag{2.8}$$

Proof. See Appendix A. □

Remark 2.5. There are several names for duality results of the form (2.8) in different areas of mathematics. Analogous results appear in optimal transport theory, see *e.g.* Villani [62], and as "perfect" or "strong" duality in the operations research literature, see *e.g.* d'Aspremont and El Ghaoui [25] and Peña et al. [52]. The superhedging duality (2.8) is crucial when verifying the ε-optimality of a measure in the numerical algorithms introduced in Section 3, see Theorem 3.24 and Corollary 3.25.

Remark 2.6. Theorem 2.4 is similar in spirit to the multi-marginal optimal transport problem, possibly under additional constraints, see *e.g.* Kellerer [45], Rachev and Rüschendorf [55], Zaev [64] and Bartl et al. [4]. However, in contrast to these articles, we do not assume that the marginals are known and fixed. Moreover, notice the subtle differences in the attainment of supremum and infimum. In Theorem 2.4, the infimum in (2.4) corresponding to the super-replication portfolio is attained, whereas the supremum on the right-hand side of (2.8) corresponding to the "worst-case" probability measure is not necessarily attained. In the multi-marginal optimal transport problem, the supremum corresponding to the "worst-case" probability measure (or optimal coupling of the marginals) is attained due to compactness, which holds when the marginal distributions of all the asset prices are fixed.

The canonical way to interpret the framework developed above is as follows: when $\Omega = \mathbb{R}^d_+$, then there exist d underlying risky assets that are traded in the financial market, and Ω represents the (non-negative) prices of the assets at a fixed future date. Investing into a unit of the asset i then corresponds to the payoff function $g(\boldsymbol{x}) \equiv \mathrm{proj}_i(\boldsymbol{x}) := x_i$ for $\boldsymbol{x} \in \mathbb{R}^d_+$. Moreover, there exist m traded derivatives (typically $m \gg d$) with known bid and ask prices $(\underline{\pi}_j, \overline{\pi}_j)_{j=1:m}$, written either on single or on multiple assets. These derivatives encode all the information available in this market. We assume that the marginal distributions of the d assets are *not* explicitly known, and we will use the information about these marginals *implied* by single-asset option prices. In particular, we consider subsets $\mathcal{J}_1, \ldots, \mathcal{J}_d \subset \{1, \ldots, m\}$ that represent the single-asset options traded for each of the d assets. Specifically, for all $j \in \mathcal{J}_i$, the derivative j has payoff function $g_j : \mathbb{R}^d_+ \mapsto \mathbb{R}$ that depends only on the asset i, *i.e.* $g_j = \widetilde{g}_j \circ \mathrm{proj}_i$ for some $\widetilde{g}_j : \mathbb{R}_+ \mapsto \mathbb{R}$. The option-implied distribution μ_i for each asset i must satisfy

$$\underline{\pi}_j \leq \int_{\mathbb{R}_+} \widetilde{g}_j \mathrm{d}\mu_i \leq \overline{\pi}_j, \quad j \in \mathcal{J}_i, \; i = 1, \ldots, d, \tag{2.9}$$

where $\underline{\pi}_j, \overline{\pi}_j$ denote the bid and ask prices. Assuming, for example, that $g_j(\boldsymbol{x}) = (x_i - \kappa_j)^+$, then inequality (2.9) means that we will consider only those marginal distributions μ_i that provide prices for a call option with strike κ_j that lie between $\underline{\pi}_j$ and $\overline{\pi}_j$.

Analogously, we assume that *partial* information on the joint distribution is available, and is again *implied* by the prices of traded multi-asset derivatives. Let $\mathcal{J} \subset \{1, \ldots, m\}$ represent the multi-asset options traded, then the option-implied joint distribution μ must satisfy

$$\underline{\pi}_j \leq \int_{\mathbb{R}^d_+} g_j \mathrm{d}\mu \leq \overline{\pi}_j, \quad j \in \mathcal{J}, \tag{2.10}$$

where again $g_j : \mathbb{R}^d_+ \mapsto \mathbb{R}$ is the payoff function of derivative j, and $\underline{\pi}_j, \overline{\pi}_j$ denote the bid and ask prices for this derivative. Assuming, for example, that $g_j(\boldsymbol{x}) = (\langle \boldsymbol{w}, \boldsymbol{x} \rangle - \kappa_j)^+$, then inequality (2.10) means that we will consider only those joint distributions μ that provide prices for basket option with weight $\boldsymbol{w}$ and strike κ_j that lie between $\underline{\pi}_j$ and $\overline{\pi}_j$. The number m of traded derivatives then equals $\sum_{i=1}^d |\mathcal{J}_i| + |\mathcal{J}|$.

In this setting, the right hand side of (2.8) is the model-free upper bound for the price of an option with payoff function f written on these d assets. The optimization takes place over all probability measures that are compatible with the option-implied information in (2.9) and (2.10), *i.e.* all probability measures that produce prices for a given option within its respective bid and ask prices. The duality result in (2.8) states that this model-free upper bound equals the least superhedging price achieved by trading in the m derivatives according to the strategy $(c, \boldsymbol{y})$, *i.e.* holding c units of cash and y_j units of derivative j for $j = 1, \ldots, m$, where the minimization takes place over all $(c, \boldsymbol{y})$ such that the payoff f is dominated, *i.e.* $c + \langle \boldsymbol{y}, \boldsymbol{g} \rangle \geq f$. The appearance of c here is due to the fact that we consider only probability measures, *i.e.* $\int_\Omega 1 \mathrm{d}\mu = 1$.

Remark 2.7. We can easily recover several frameworks existing in the literature by an appropriate choice of the function $\boldsymbol{g}$. Indeed, d'Aspremont and El Ghaoui [25], Peña et al. [51, 52] and Cho, Kim, and Lee [22] compute model-free bounds for basket options assuming that prices of vanilla call options, forward prices, and prices of basket options are available in the market.

The next example demonstrates that the supremum on the right-hand side of (2.8) is not necessarily attained.

Example 2.8. Let $\Omega = \mathbb{R}_+$, $m = 1$, and, for $x \in \mathbb{R}_+$, set:

$$g_1(x) = x, \quad f(x) = \max\{x - 1, 0\}, \quad \underline{\pi}_1 = 0 \text{ and } \overline{\pi}_1 = 1.$$

Clearly,

$$\begin{aligned} & c + y_1 g_1(x) - f(x) \geq 0 \quad \forall x \in \mathbb{R}_+ \\ \Longleftrightarrow \quad & c + y_1 x - \max\{x - 1, 0\} \geq 0 \quad \forall x \in \mathbb{R}_+ \\ \Longrightarrow \quad & c \geq 0, \ y_1 \geq 1. \end{aligned}$$

In case $c = 0$, $y_1 = 1$, then $c + y_1 g_1(x) - f(x) \geq 0 \ \forall x \in \mathbb{R}_+$ holds, and we have that $c^\star = 0$, $y_1^\star = 1$ is an optimizer of (2.4). Thus $\phi(f) = c^\star + y_1^\star \overline{\pi}_1 = 1$. On the other hand, if μ is a probability measure on $\mathbb{R}_+$, $\underline{\pi}_1 \leq \int_{\mathbb{R}_+} g_1 \mathrm{d}\mu \leq \overline{\pi}_1$, and $\int_{\mathbb{R}_+} f \mathrm{d}\mu = 1$, then

$$\int_{\mathbb{R}_+} \min\{x, 1\} \mathrm{d}\mu = \int_{\mathbb{R}_+} (g_1 - f) \mathrm{d}\mu = \int_{\mathbb{R}_+} g_1 \mathrm{d}\mu - \int_{\mathbb{R}_+} f \mathrm{d}\mu \leq \overline{\pi}_1 - 1 = 0.$$

Since $\min\{x, 1\} > 0$ for $x > 0$, it is implied that $\mu((0, +\infty)) = 0$, which is impossible. Hence, $\sup_{\mu \in \mathcal{Q}} \int_\Omega f \mathrm{d}\mu$ is not attained.

The next proposition presents a specific setting in which Assumption 2.1 holds. This setting will be used in the numerical experiments in Section 4.

Proposition 2.9. *Let $\Omega \subset \mathbb{R}^d$ for some $d \in \mathbb{N}$, $g_1, \ldots, g_m$ be continuous functions on Ω, and assume there exists $\widehat{\mu} \in \mathcal{Q}$ such that $\widehat{\mu}$ is equivalent to the Lebesgue measure on Ω. Then, Assumption 2.1 holds.*

Proof. See Appendix A. □

3. NUMERICAL METHODS FOR THE COMPUTATION OF BOUNDS

The superhedging duality in Theorem 2.4 allows to transform the abstract maximization problem over probability measures, into a more tractable minimization problem over vectors that satisfy certain constraints. The aim of this section is to develop novel numerical methods for the exact and efficient computation of upper and lower bounds on $\phi(f)$. More specifically, we will develop methods for the computation of upper and lower bounds $\phi(f)^{\mathrm{UB}}$ and $\phi(f)^{\mathrm{LB}}$ which are ε-optimal, *i.e.*

$$\phi(f)^{\mathrm{LB}} \leq \phi(f) \leq \phi(f)^{\mathrm{UB}} \quad \text{and} \quad \phi(f)^{\mathrm{UB}} - \phi(f)^{\mathrm{LB}} \leq \varepsilon,$$

for $\varepsilon > 0$. Our methods allow us to also characterize the optimal pricing measure associated with the primal maximization problem. Therefore, we provide a complete solution to both optimization problems, and can characterize the solution both in terms of ε-optimal hedging strategies and in terms of the optimal pricing measure.

Let $\overline{\boldsymbol{\pi}} := (\overline{\pi}_1, \dots, \overline{\pi}_m)^\mathsf{T}$ and $\underline{\boldsymbol{\pi}} := (\underline{\pi}_1, \dots, \underline{\pi}_m)^\mathsf{T}$. The minimization problem $\phi(f)$ in (2.4) can be equivalently formulated as a linear semi-infinite programming (LSIP) problem, *i.e.* as an optimization problem with a linear objective and an infinite number of linear constraints, one for each $\omega \in \Omega$,

$$\begin{aligned} \phi(f) = \quad & \text{minimize} \quad c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}} \rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}} \rangle \\ & \text{subject to} \quad c + \langle \boldsymbol{y}^+ - \boldsymbol{y}^-, \boldsymbol{g}(\omega) \rangle \geq f(\omega), \forall \omega \in \Omega, \\ & \qquad\qquad\quad c \in \mathbb{R}, \boldsymbol{y}^+ \geq \mathbf{0}, \boldsymbol{y}^- \geq \mathbf{0}. \end{aligned} \tag{3.1}$$

LSIP problems are classical optimization problems that have been thoroughly studied in the related literature; see *e.g.* Goberna and López [37, 38] and Reemtsen and Rückmann [56]. Many numerical methods exist for solving LSIP problems under certain regularity assumptions. In this section, we develop novel algorithms tailored to solving (3.1) under different assumptions on Ω, $\boldsymbol{g}$ and f.

In the following subsections, we consider two settings where we work with *continuous piece-wise affine* (CPWA) payoff functions. In the first setting, we assume that $\Omega = \mathbb{R}^d_+$ (*i.e.* non-negative asset prices), and we develop an algorithm that is based on the cutting plane discretization method (see Goberna and López [37], Algorithm 11.4.1), which we refer to as the *exterior cutting plane* (ECP) method. In the second setting, we assume that $\Omega = [0, \overline{x}_1] \times [0, \overline{x}_2] \times \dots \times [0, \overline{x}_d]$ for some $\overline{x}_1, \overline{x}_2, \dots, \overline{x}_d > 0$ (*i.e.* non-negative and upper-bounded asset prices). Under this more restrictive assumption, we develop an algorithm that is based on the accelerated central cutting plane (ACCP) method by Betrò [11].

Suppose now that $\phi(f) < \infty$. Let

$$S := \left\{ \left(\begin{smallmatrix} c \\ \boldsymbol{y} \end{smallmatrix} \right) \in \mathbb{R}^{m+1} : c + \langle \boldsymbol{y}, \boldsymbol{g} \rangle \geq f \right\} \tag{3.2}$$

denote the set of portfolios that super-replicate f; this set is closed and convex. Let

$$S^\star := \left\{ \left(\begin{smallmatrix} c \\ \boldsymbol{y} \end{smallmatrix} \right) \in S : c + \pi(\boldsymbol{y}) = \phi(f) \right\} \tag{3.3}$$

denote the set of minimizers of (2.4), which is again closed and convex. Under Assumption 2.1 and the assumption that $\phi(f) < \infty$, $S^\star$ is non-empty by Theorem 2.4(ii).

Before introducing the settings, let us first introduce the notion of CPWA functions and some examples from finance.

Definition 3.1 (Continuous piece-wise affine function)**.** We call a function $h : \mathbb{R}^d \mapsto \mathbb{R}$ a CPWA function if it can be represented as

$$h(\boldsymbol{x}) = \sum_{k=1}^{K} \xi_k \max \left\{ \langle \boldsymbol{a}_{k,i}, \boldsymbol{x} \rangle + b_{k,i} : 1 \leq i \leq I_k \right\}, \tag{3.4}$$

where $K \in \mathbb{N}$, $I_k \in \mathbb{N}$ for $k = 1, \dots, K$, and $\boldsymbol{a}_{k,i} \in \mathbb{R}^d$, $b_{k,i} \in \mathbb{R}$, $\xi_k \in \{-1, 1\}$ for $i = 1, \dots, I_k, k = 1, \dots, K$.

Example 3.2. Many popular payoff functions in finance belong to the class of CPWA functions. The list below contains those payoff functions used in the present work, alongside their CPWA representation.

(i) Setting $K = 1, \xi_1 = 1, I_1 = 2, \boldsymbol{a}_{1,1} = \boldsymbol{e}_i, \boldsymbol{a}_{1,2} = \mathbf{0}$ and $b_{1,1} = -\kappa, b_{1,2} = 0$, we get the payoff of a call option on the i-th asset with strike κ, via

$$h(\boldsymbol{x}) = \max\{x_i - \kappa, 0\} = (x_i - \kappa)^+.$$

(ii) Using the previous setting and replacing $\boldsymbol{a}_{1,1}$ with $\boldsymbol{a}_{1,1} = \boldsymbol{w} = (w_1, \dots, w_d)^\mathsf{T} \in \mathbb{R}^d_+$, we get the payoff of a basket call option on the d assets with weights $\boldsymbol{w}$ and strike κ, via

$$h(\boldsymbol{x}) = \max \left\{ \sum_{i=1}^{d} w_i x_i - \kappa, 0 \right\} = \left(\sum_{i=1}^{d} w_i x_i - \kappa \right)^+.$$

(iii) Using the previous setting, and replacing $\boldsymbol{a}_{1,1}$ with $\boldsymbol{a}_{1,1} = \sum_{i\in\mathcal{I}} \boldsymbol{e}_i - \sum_{j\in\mathcal{I}'} \boldsymbol{e}_j$, for $\mathcal{I}, \mathcal{I}' \subset \{1,\dots,d\}$ and $\mathcal{I}\cap\mathcal{I}' = \emptyset$, we get the payoff of a spread call option with strike κ, via

$$h(\boldsymbol{x}) = \max\Big\{\sum_{i\in\mathcal{I}} x_i - \sum_{j\in\mathcal{I}'} x_j - \kappa, 0\Big\} = \Big(\sum_{i\in\mathcal{I}} x_i - \sum_{j\in\mathcal{I}'} x_j - \kappa\Big)^+.$$

(iv) Setting $K = 1, \xi_1 = 1, I_1 = d+1, \boldsymbol{a}_{1,i} = \boldsymbol{e}_i, b_{1,i} = -\kappa$ for all $i \in \{1,\dots,d\}$, $\boldsymbol{a}_{1,d+1} = \boldsymbol{0}$ and $b_{1,d+1} = 0$, we get the payoff of a call-on-max option on the d assets with strike κ, via

$$\begin{aligned} h(\boldsymbol{x}) &= \max\big\{x_1 - \kappa, \dots, x_d - \kappa, 0\big\} = \max\big\{\max\big\{x_1,\dots,x_d\big\} - \kappa, 0\big\} \\ &= \big(x_1 \vee \dots \vee x_d - \kappa\big)^+. \end{aligned}$$

(v) Setting $K = 2, \xi_1 = 1, \xi_2 = -1, I_1 = d+1, I_2 = d, \boldsymbol{a}_{1,i} = \boldsymbol{a}_{2,i} = -\boldsymbol{e}_i, b_{1,i} = b_{2,i} = \kappa$ for all $i \in \{1,\dots,d\}$, and $\boldsymbol{a}_{1,d+1} = \boldsymbol{0}$, $b_{1,d+1} = 0$, we get the payoff of a call-on-min option on the d assets with strike κ, via

$$\begin{aligned} h(\boldsymbol{x}) &= \max\big\{\kappa - x_1, \dots, \kappa - x_d, 0\big\} - \max\big\{\kappa - x_1, \dots, \kappa - x_d\big\} \\ &= \max\big\{\min\big\{\kappa - x_1,\dots,\kappa - x_d\big\}, 0\big\} \\ &= \big(x_1 \wedge \dots \wedge x_d - \kappa\big)^+. \end{aligned}$$

(vi) Finally, setting $K = 1, \xi_1 = 1, I_1 = d+1, \boldsymbol{a}_{1,i} = \boldsymbol{e}_i, b_{1,i} = -\kappa_i$ for all $i \in \{1,\dots,d\}$, and $\boldsymbol{a}_{1,d+1} = \boldsymbol{0}, b_{1,d+1} = 0$, we get the payoff of a best-of-call option on the d assets with strikes $\kappa_1,\dots,\kappa_d$, via

$$\begin{aligned} h(\boldsymbol{x}) &= \max\big\{x_1 - \kappa_1, \dots, x_d - \kappa_d, 0\big\} \\ &= \max\big\{(x_1 - \kappa_1)^+, \dots, (x_d - \kappa_d)^+\big\} \\ &= (x_1 - \kappa_1)^+ \vee (x_2 - \kappa_2)^+ \vee \dots \vee (x_d - \kappa_d)^+. \end{aligned}$$

Let us point out that these representations are not unique. Moreover, by replacing the vectors $\boldsymbol{e}_i$ or $-\boldsymbol{e}_i$ with suitable vectors in examples (iii), (iv), (v), and (vi) above, one can create weighted versions of the aforementioned payoff. These can be interpreted as options written on a number of indexes.

Given Definition 3.1, the following properties of CPWA functions can be deduced.

Lemma 3.3. *The following are properties of CPWA functions:*

(i) A finite linear combination of CPWA functions is again a CPWA function;

(ii) If $h : \mathbb{R}^d_+ \mapsto \mathbb{R}$ is a CPWA function, then it admits the following local representation:

$$h(\boldsymbol{x}) = \begin{cases} \langle \boldsymbol{a}_1, \boldsymbol{x}\rangle + b_1, & \text{if } \boldsymbol{x} \in \Omega_1, \\ \quad\vdots & \quad\vdots \\ \langle \boldsymbol{a}_J, \boldsymbol{x}\rangle + b_J, & \text{if } \boldsymbol{x} \in \Omega_J, \end{cases} \tag{3.5}$$

where for $j = 1,\dots,J$, $\boldsymbol{a}_j \in \mathbb{R}^d$, $b_j \in \mathbb{R}$, Ω_j is a closed polyhedron, and $\bigcup_{j=1}^J \Omega_j = \mathbb{R}^d_+$. In addition, for $j \neq j'$, $\operatorname{int}(\Omega_j)\cap\operatorname{int}(\Omega_{j'}) = \emptyset$. If $\boldsymbol{x} \in \Omega_j\cap\Omega_{j'}$, then $\langle \boldsymbol{a}_j, \boldsymbol{x}\rangle + b_j = \langle \boldsymbol{a}_{j'}, \boldsymbol{x}\rangle + b_{j'}$, and thus the above function is well-defined.

(iii) In the local representation (3.5), it can be assumed, without loss of generality, that $\operatorname{int}(\Omega_j) \neq \emptyset$ for $j = 1,\dots,J$.

Proof. See Appendix B. □

For any CPWA function defined on $\mathbb{R}^d_+$, we introduce its radial function, which is defined as follows.

Definition 3.4 (Radial function of a CPWA function)**.** Let $h : \mathbb{R}^d_+ \mapsto \mathbb{R}$ be a CPWA function with representation (3.4). The *radial function* of h, denoted by $\widetilde{h} : \mathbb{R}^d_+ \mapsto \mathbb{R}$, is defined as

$$\widetilde{h}(\boldsymbol{z}) := \sum_{k=1}^K \xi_k \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{z}\rangle : 1 \le i \le I_k\right\}. \tag{3.6}$$

The following proposition establishes the connection between the lower-boundedness of a CPWA function and the non-negativity of its associated radial function.

Proposition 3.5. *Let $h : \mathbb{R}^d_+ \mapsto \mathbb{R}$ be a CPWA function, and let $\widetilde{h} : \mathbb{R}^d_+ \mapsto \mathbb{R}$ be its radial function. Then,*

(i) $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} h(\boldsymbol{x}) > -\infty$ *if and only if* $\widetilde{h}(\boldsymbol{z}) \geq 0$ *for all* $\boldsymbol{z} \in \mathbb{R}^d_+$*;*

(ii) if $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} h(\boldsymbol{x}) > -\infty$*, there exists* $\boldsymbol{x}^\star \in \mathbb{R}^d_+$ *such that* $h(\boldsymbol{x}^\star) = \inf_{\boldsymbol{x}\in\mathbb{R}^d_+} h(\boldsymbol{x})$*.*

Proof. See Appendix B. □

3.1. **CPWA payoff functions on unbounded domains.** In the first setting, we work under the following assumptions.

Assumption 3.6 (Setting 1)**.** We assume the following:

(i) $\Omega = \mathbb{R}^d_+$;

(ii) f and $(g_j)_{j=1:m}$ are CPWA functions on Ω;

(iii) $\phi(f) < \infty$ and $\phi(-f) < \infty$.

In the sequel, for notational reasons, we use $\boldsymbol{x}$ in place of ω when Ω is a subset of the Euclidean space. Let us now introduce the notion of the slack function for the LSIP problem in (3.1).

Definition 3.7 (Slack function)**.** Let $\boldsymbol{y} \in \mathbb{R}^m$ be fixed, and denote the slack function of the LSIP problem in (3.1) by $s_{\boldsymbol{y}} : \mathbb{R}^d_+ \mapsto \mathbb{R}$. This is defined as follows:

$$s_{\boldsymbol{y}}(\boldsymbol{x}) := \langle \boldsymbol{y}, \boldsymbol{g}(\boldsymbol{x})\rangle - f(\boldsymbol{x}). \tag{3.7}$$

Remark 3.8. By Definition 3.7, the semi-infinite constraint in (3.1)

$$c + \langle \boldsymbol{y}^+ - \boldsymbol{y}^-, \boldsymbol{g}(\boldsymbol{x})\rangle \geq f(\boldsymbol{x}), \forall \boldsymbol{x} \in \mathbb{R}^d_+ \quad \iff \quad \inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}^+-\boldsymbol{y}^-}(\boldsymbol{x}) \geq -c.$$

Definition 3.7 is slightly different from the usual notion of slack function, as the usual definition also includes c. We choose to define the slack function without c for notational simplicity.

Given Definition 3.7 and Assumption 3.6, we arrive at the following results using Lemma 3.3 and Proposition 3.5.

Proposition 3.9. *Under Assumption 3.6, the following statements hold:*

(i) For any fixed $\boldsymbol{y} \in \mathbb{R}^m$*, the slack function* $s_{\boldsymbol{y}} : \mathbb{R}^d_+ \mapsto \mathbb{R}$ *is a CPWA function. For any fixed* $\boldsymbol{x} \in \mathbb{R}^d_+$*,* $s_{\boldsymbol{y}}(\boldsymbol{x})$ *is linear when regarded as a function of* $\boldsymbol{y}$*. In particular,* $s_{\boldsymbol{y}}(\boldsymbol{x}) \geq -c$ *corresponds to a linear inequality constraint on* $(c, \boldsymbol{y})$ *for every fixed* $\boldsymbol{x} \in \mathbb{R}^d_+$*.*

(ii) There exist $K \in \mathbb{N}$*,* $\boldsymbol{w}_k \in \mathbb{R}^d$*,* $z_k \in \mathbb{R}$*,* $I_k \in \mathbb{N}$ *for* $k = 1, \ldots, K$*,* $\boldsymbol{a}_{k,i} \in \mathbb{R}^d$*,* $b_{k,i} \in \mathbb{R}$*, for* $i = 1, \ldots, I_k, k = 1, \ldots, K$*, such that*

$$s_{\boldsymbol{y}}(\boldsymbol{x}) = \sum_{k=1}^{K} (\langle \boldsymbol{y}, \boldsymbol{w}_k\rangle + z_k) \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}\rangle + b_{k,i} : 1 \leq i \leq I_k\right\}. \tag{3.8}$$

Let the radial function of $s_{\boldsymbol{y}}$ *be denoted by* $\widetilde{s}_{\boldsymbol{y}}$*. There exist* $\widetilde{K} \in \mathbb{N}$*,* $\widetilde{\boldsymbol{w}}_k \in \mathbb{R}^d$*,* $\widetilde{z}_k \in \mathbb{R}$*,* $\widetilde{I}_k \in \mathbb{N}$ *for* $k = 1, \ldots, K$*,* $\widetilde{\boldsymbol{a}}_{k,i} \in \mathbb{R}^d$*, for* $i = 1, \ldots, \widetilde{I}_k, k = 1, \ldots, \widetilde{K}$*, such that*

$$\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) = \sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k) \max\left\{\langle \widetilde{\boldsymbol{a}}_{k,i}, \boldsymbol{z}\rangle : 1 \leq i \leq \widetilde{I}_k\right\}. \tag{3.9}$$

(iii) The slack function $s_{\boldsymbol{y}}(\boldsymbol{x})$ *is bounded from below as a function of* $\boldsymbol{x}$*. Moreover,* $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x})$ *is attained at some* $\boldsymbol{x}^\star$ *if and only if* $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0$ *for all* $\boldsymbol{z} \in \mathbb{R}^d_+$*.*

(iv) There exists a finite set $X^\star \subset \mathbb{R}^d_+$ *such that for every* $\boldsymbol{y} \in \mathbb{R}^m$ *that satisfies* $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x}) > -\infty$*, there exists* $\boldsymbol{x}^\star \in X^\star$ *such that* $s_{\boldsymbol{y}}(\boldsymbol{x}^\star) = \inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x})$*.*

Proof. See Appendix B.1.1. □

A crucial step in the exterior cutting plane algorithm (Line 12 of Algorithm 1) that will be introduced later is to minimize the slack function $s_{\boldsymbol{y}}(\boldsymbol{x})$ over $\boldsymbol{x} \in \mathbb{R}^d_+$. In order to do so, one needs to first guarantee that $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x}) > -\infty$, which, by Proposition 3.9(iii), is equivalent to $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0$ for all $\boldsymbol{z} \in \mathbb{R}^d_+$. The next proposition details necessary and sufficient conditions on $\boldsymbol{y}$ such that $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0$ for all $\boldsymbol{z} \in \mathbb{R}^d_+$, and leads to the computational approach to generate such constraints.

Proposition 3.10. *Let $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z})$ be the slack function as in Proposition 3.9(ii). Let $\widetilde{\mathcal{I}} = \{(i_k)_{k=1:\widetilde{K}} : 1 \leq i_k \leq \widetilde{I}_k, k = 1,\dots,\widetilde{K}\}$. For $(i_k) \in \widetilde{\mathcal{I}}$, define $A_{(i_k)} := \{\widetilde{\boldsymbol{a}}_{k,i_k} - \widetilde{\boldsymbol{a}}_{k,i} : i \in \{1,\dots,\widetilde{I}_k\} \setminus \{i_k\}, k = 1,\dots,\widetilde{K}\}$. Let $\operatorname{dual}(C)$ denote the dual cone of a set $C \subset \mathbb{R}^d$. Then, the following statements hold:*

(i) For each $(i_k) \in \widetilde{\mathcal{I}}$, it holds that

$$\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) = \sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k)\langle \widetilde{\boldsymbol{a}}_{k,i_k}, \boldsymbol{z}\rangle,$$

for all $\boldsymbol{z} \in \operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+$. We have $\bigcup_{(i_k)\in\widetilde{\mathcal{I}}} \operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+ = \mathbb{R}^d_+$. If $(i_k) \in \widetilde{\mathcal{I}}$, $(i'_k) \in \widetilde{\mathcal{I}}$, $(i_k) \neq (i'_k)$, then

$$\operatorname{int}(\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) \cap \operatorname{int}(\operatorname{dual}(A_{(i'_k)}) \cap \mathbb{R}^d_+) = \emptyset.$$

Moreover, by possibly removing some elements from $\widetilde{\mathcal{I}}$, we can assume, without loss of generality, that $\operatorname{int}(\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) \neq \emptyset$ for all $(i_k) \in \widetilde{\mathcal{I}}$.

(ii) For each $(i_k) \in \widetilde{\mathcal{I}}$, $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0$ for all $\boldsymbol{z} \in \operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+$ if and only if for each $\boldsymbol{v} \in A_{(i_k)}$ there exists $\eta^{(i_k)}_{\boldsymbol{v}} \geq 0$ such that

$$\sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k)\widetilde{\boldsymbol{a}}_{k,i_k} \geq \sum_{\boldsymbol{v}\in A_{(i_k)}} \eta^{(i_k)}_{\boldsymbol{v}} \boldsymbol{v}. \tag{3.10}$$

(iii) For each $(i_k) \in \widetilde{\mathcal{I}}$, $\operatorname{int}(\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) = \emptyset$ is equivalent to the following condition:

$$\begin{aligned} \text{there exists} \quad & \eta^{(i_k)}_{\boldsymbol{v}} \geq 0 \text{ for each } \boldsymbol{v} \in A_{(i_k)}, \\ \text{such that} \quad & \sum_{\boldsymbol{v}\in A_{(i_k)}} \eta^{(i_k)}_{\boldsymbol{v}} = 1, \\ & \sum_{\boldsymbol{v}\in A_{(i_k)}} \eta^{(i_k)}_{\boldsymbol{v}} \boldsymbol{v} \leq \boldsymbol{0}, \end{aligned} \tag{3.11}$$

which can be checked by a linear programming (LP) solver.

Proof. See Appendix B.1.1. □

Based on Proposition 3.10, Algorithm 0 generates the necessary and sufficient constraints on $\boldsymbol{y}$ to guarantee that $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0$ for all $\boldsymbol{z} \in \mathbb{R}^d_+$, which are jointly referred to as *radial constraints*. The output of Algorithm 0 is a system of linear inequalities of the form (3.10) and is denoted by $\widetilde{\sigma}$. Apart from $\boldsymbol{y}$, these linear inequalities also contain auxiliary variables. In Line 5, the system of linear inequalities is "expanded" to include additional auxiliary variables. In Line 6, a new linear inequality is added to the current inequality system. This is sometimes referred to as *aggregating* the constraint to $\widetilde{\sigma}$ in the literature. Concretely, if the linear inequality constraints are stored in a matrix, then Line 5 corresponds to adding extra columns to the matrix to increase the number of decision variables, and Line 6 corresponds to adding an extra row to the matrix as an additional constraint.

Algorithm 0: Generation of Radial Constraints

Input: $\boldsymbol{y} \in \mathbb{R}^m$, $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) = \sum_{k=1}^{\widetilde{K}}(\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k \rangle + \widetilde{z}_k) \max \left\{ \langle \widetilde{\boldsymbol{a}}_{k,i}, \boldsymbol{z} \rangle : 1 \leq i \leq \widetilde{I}_k \right\}$ for all $\boldsymbol{z} \in \mathbb{R}^d_+$

Output: System of linear inequalities on $\boldsymbol{y}$ with auxiliary variables to guarantee $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0$ for all $\boldsymbol{z} \in \mathbb{R}^d_+$, denoted by $\widetilde{\sigma}$

1 Initialize $\widetilde{\sigma}$ to be empty.
2 **for** *every* $(i_k) \in \widetilde{\mathcal{I}}$ **do**
3   Compute $A_{(i_k)} \leftarrow \{\widetilde{\boldsymbol{a}}_{k,i_k} - \widetilde{\boldsymbol{a}}_{k,i} : 1 \leq i \leq \widetilde{I}_k, i \neq i_k, 1 \leq k \leq \widetilde{K}\}$.
4   **if** *the condition (3.11) holds* **then**
5   Expand $\widetilde{\sigma}$ to include additional non-negative auxiliary variables $(\eta_{\boldsymbol{v}}^{(i_k)})_{\boldsymbol{v} \in A_{(i_k)}}$.
6   Append the linear inequality (3.10) to $\widetilde{\sigma}$.
7 **return** $\widetilde{\sigma}$.

Remark 3.11. (i) In Algorithm 0, auxiliary variables are introduced because each $\operatorname{cone}(A_{(i_k)})$ is represented by its extreme directions $A_{(i_k)}$ (*i.e.* a V-polyhedron). Such a cone also admits a representation by faces (*i.e.* H-polyhedron), that is, $\{\boldsymbol{v} \in \mathbb{R}^d_+ : \langle \boldsymbol{v}, \boldsymbol{u} \rangle \geq 0, \forall \boldsymbol{u} \in \operatorname{exdir}(\operatorname{dual}(\operatorname{cone}(A_{(i_k)})))\}$, where $\operatorname{exdir}(\cdot)$ denotes the set of extreme directions of a polyhedron. Using this representation, there is no need to introduce auxiliary variables. Unfortunately, to obtain such a representation one needs to enumerate all the extreme directions of $\operatorname{dual}(\operatorname{cone}(A_{(i_k)}))$, which is usually computationally very costly (see, for example, Le Verge [46] and Section 5.2 of Ciripoi, Löhne, and Weißing [23]).

(ii) The computational complexity of Algorithm 0 may grow fast as $\widetilde{K}$ grows. In certain cases, the growth can be exponential. The proposed approach is only practical when the total number of times Line 6 is executed is not too large. Fortunately, this is usually the case with the problems we study, since vanilla options and basket options with non-negative weights do not increase this number.

Now that we have generated the radial constraints, for each $\boldsymbol{y}$ satisfying all linear inequality constraints in $\widetilde{\sigma}$, we have $\inf_{\boldsymbol{x} \in \mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x}) > -\infty$. By Proposition 3.9(iii), the minimizer is attained. Our next goal is to be able to solve the global minimization problem $\min_{\boldsymbol{x} \in \mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x})$ for fixed $\boldsymbol{y}$. We have the following lemma that can help us solve the problem.

Lemma 3.12. *Let $h : \mathbb{R}^d_+ \mapsto \mathbb{R}$ be a CPWA function as in* (3.4). *Assume that $\inf_{\boldsymbol{x} \in \mathbb{R}^d_+} h(\boldsymbol{x}) > -\infty$. Let $\overline{\boldsymbol{x}} := (\overline{x}_1, \ldots, \overline{x}_d)^\mathsf{T} > \boldsymbol{0}$. For $i = 1, \ldots, I_k$, $k = 1, \ldots, K$, let $M_{k,i}$ be the constant given by*

$$M_{k,i} := \max\{\langle \boldsymbol{a}_{k,i'} - \boldsymbol{a}_{k,i}, \boldsymbol{x} \rangle + b_{k,i'} - b_{k,i} : \boldsymbol{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}, 1 \leq i' \leq I_k, i' \neq i\}. \tag{3.12}$$

Then, the global minimization problem

$$\begin{aligned} \textit{minimize} \quad & h(\boldsymbol{x}) = \sum_{k=1}^{K} \xi_k \max\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x} \rangle + b_{k,i} : 1 \leq i \leq I_k\} \\ \textit{subject to} \quad & \boldsymbol{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}} \end{aligned} \tag{3.13}$$

is equivalent to the following mixed-integer linear programming (MILP) problem:

$$
\begin{aligned}
\textit{minimize} \quad & \sum_{k=1,\dots,K,\xi_k=1} \lambda_k + \sum_{k=1,\dots,K,\xi_k=-1} -\zeta_k \\
\textit{subject to} \quad & \textit{for } k=1,\dots,K \textit{ and } \xi_k = 1: \\
& \qquad \begin{cases} \lambda_k \in \mathbb{R}, \\ \langle \boldsymbol{a}_{k,i}, \boldsymbol{x} \rangle + b_{k,i} \leq \lambda_k, \quad \forall 1 \leq i \leq I_k, \end{cases} \\
& \textit{for } k=1,\dots,K \textit{ and } \xi_k = -1: \\
& \qquad \begin{cases} \zeta_k \in \mathbb{R}, \\ \delta_{k,i} \in \mathbb{R}, & \forall 1 \leq i \leq I_k, \\ \iota_{k,i} \in \{0,1\} & \forall 1 \leq i \leq I_k, \\ \langle \boldsymbol{a}_{k,i}, \boldsymbol{x} \rangle + b_{k,i} + \delta_{k,i} = \zeta_k & \forall 1 \leq i \leq I_k, \\ 0 \leq \delta_{k,i} \leq M_{k,i}(1-\iota_{k,i}) & \forall 1 \leq i \leq I_k, \end{cases} \\
& \qquad \sum_{i=1}^{I_k} \iota_{k,i} = 1, \\
& \boldsymbol{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}.
\end{aligned}
\tag{3.14}
$$

Proof. See Appendix B.1.1. □

The MILP problem in (3.14) can be solved efficiently by state-of-the-art solvers such as Gurobi [39] that uses the so-called Branch-and-Bound (BnB) algorithm.

Remark 3.13. Consider the MILP problem: minimize $p(\boldsymbol{\gamma})$ subject to $\boldsymbol{\gamma} \in S$, where S denotes the feasible set. The BnB algorithm generates a sequence of solutions $(\boldsymbol{\gamma}_t) \subset S$ that satisfy the equality/inequality constraints and integral constraints with non-increasing objective values, that is, for $\overline{p}_t := p(\boldsymbol{\gamma}_t)$, we have

$$\overline{p}_1 \geq \overline{p}_2 \geq \dots \geq p^\star,$$

where $p^\star$ denotes the optimal value of the problem. These solutions are usually referred to as integer feasible solutions. At the same time, the BnB algorithm produces a sequence of relaxations of the original problem with non-decreasing objective values to serve as lower bounds of the original problem, that is,

$$\underline{p}_1 \leq \underline{p}_2 \leq \dots \leq p^\star.$$

The BnB algorithm terminates on iteration t when the following condition is satisfied:

$$\frac{\overline{p}_t - \underline{p}_t}{|\overline{p}_t|} \leq \zeta, \tag{3.15}$$

where $\zeta \in (0,1)$ is referred to as the relative gap tolerance. At termination, $\boldsymbol{\gamma}_t$ is returned as the approximate optimizer, $\overline{p}_t$ is returned as the approximate optimal value, and $\underline{p}_t$ is returned as the best objective bound. Gurobi Optimization provides a basic overview of mixed-integer programming[1].

We are now ready to introduce the complete implementation of the cutting plane discretization method, shown in Algorithm 1. We will refer to this method as the exterior cutting plane (ECP) method, since every constraint generated in Line 8 of Algorithm 1 (also known as cut) does not restrict the feasible set S defined in (3.2), hence is exterior to S. In the following, Remark 3.14 explains the inputs to Algorithm 1, Remark 3.15 explains some of its unique features, and Theorem 3.16 shows its validity.

[1] https://www.gurobi.com/resource/mip-basics/

Algorithm 1: Exterior Cutting Plane (ECP) Algorithm (under Assumption 3.6)

Input: $\overline{\boldsymbol{\pi}}$, $\underline{\boldsymbol{\pi}}$, $(g_j)_{j=1:m}$, f, $X^{(0)} \subset \mathbb{R}^d_+$, $\underline{\phi}$, $\overline{\boldsymbol{x}} > \mathbf{0}$, $\varepsilon > 0$, $\tau > 0$, $0 < \delta \leq 1$

Output: $\phi(f)^{\mathrm{UB}}$, $\phi(f)^{\mathrm{LB}}$, $c^\star$, $\boldsymbol{y}^\star$, X

1 Compute the relevant terms in $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) = \sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k \rangle + \widetilde{z}_k) \max \left\{ \langle \widetilde{\boldsymbol{a}}_{k,i}, \boldsymbol{z} \rangle : 1 \leq i \leq \widetilde{I}_k \right\}$.

2 Use Algorithm 0 to generate radial constraints, denoted by the system of linear inequalities with auxiliary variables $\widetilde{\sigma}$.

3 Construct a system of linear inequalities σ which contains all variables and inequalities in $\widetilde{\sigma}$, additional variables $c \in \mathbb{R}$, $\boldsymbol{y}^+ \geq \mathbf{0}$ and $\boldsymbol{y}^- \geq \mathbf{0}$, and an additional equality $\boldsymbol{y} = \boldsymbol{y}^+ - \boldsymbol{y}^-$.

4 Add the linear inequality $c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}} \rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}} \rangle \geq \underline{\phi} - \tau$ to σ.

5 $r \leftarrow 0$.

6 **repeat**

7   **for** *each* $\boldsymbol{x} \in X^{(r)}$ **do**

8   Add the linear inequality $c + \langle \boldsymbol{y}^+ - \boldsymbol{y}^-, \boldsymbol{g}(\boldsymbol{x}) \rangle \geq f(\boldsymbol{x})$ to σ.

9   Solve the LP problem: $\underline{\varphi}^{(r)} \leftarrow$ minimize $c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}} \rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}} \rangle$ subject to linear constraints σ, and denote the computed minimizer as $(c^{(r)}, \boldsymbol{y}^{+(r)}, \boldsymbol{y}^{-(r)})$.

10     $\boldsymbol{y}^{(r)} \leftarrow \boldsymbol{y}^{+(r)} - \boldsymbol{y}^{-(r)}$.

11   Formulate the global minimization problem: minimize $s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x})$ subject to $\mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}$ into the MILP problem in (3.14).

12     $s^{(r)} \leftarrow c^{(r)} + \min_{\mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}} s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x})$ (solve the MILP problem).

13     $X^{(r)} \leftarrow \Big\{ \boldsymbol{x} : (\boldsymbol{x}, (\lambda_k), (\zeta_k), (\delta_{k,i}), (\iota_{k,i}))$ is an integer feasible solution found by the BnB algorithm while solving (3.14) such that $c^{(r)} + s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x}) \leq \delta s^{(r)} \Big\}$.

14     $r \leftarrow r + 1$.

until $s^{(r-1)} \geq -\varepsilon$.

15 $\phi(f)^{\mathrm{LB}} \leftarrow \underline{\varphi}^{(r-1)}$, $\phi(f)^{\mathrm{UB}} \leftarrow \underline{\varphi}^{(r-1)} - s^{(r-1)}$, $c^\star \leftarrow c^{(r-1)} - s^{(r-1)}$, $\boldsymbol{y}^\star \leftarrow \boldsymbol{y}^{(r-1)}$, $X \leftarrow \bigcup_{l=0}^{r-1} X^{(l)}$.

16 **if** $\phi(f)^{\mathrm{UB}} < \underline{\phi}$ **then**

17   **return** The problem (3.1) is unbounded.

18 **else**

19   **return** $\phi(f)^{\mathrm{UB}}$, $\phi(f)^{\mathrm{LB}}$, $c^\star$, $\boldsymbol{y}^\star$, X.

Remark 3.14 (Inputs to Algorithm 1)**.** The following list explains the various inputs to Algorithm 1 and how to set them.

- $\overline{\boldsymbol{\pi}}$, $\underline{\boldsymbol{\pi}}$, $(g_j)_{j=1:m}$, f are given by problem (3.1).
- $X^{(0)}$ is the initial set of points in $\mathbb{R}^d_+$ used to generate the initial constraints on c and $\boldsymbol{y}$. It is possible to set $X^{(0)} = \emptyset$ and the LP problem in Line 9 remains bounded due to the additional constraint introduced in Line 4. The return value X (or a subset of X) can be reused as $X^{(0)}$ for another run of Algorithm 1 with a similar problem to achieve significant speed-up.
- $\overline{\boldsymbol{x}}$ is the upper bound used in the MILP problem (3.14). For the validity of the algorithm, $\overline{\boldsymbol{x}}$ needs to be large enough such that $\boldsymbol{x}^\star \leq \overline{\boldsymbol{x}}$ for all $\boldsymbol{x}^\star \in X^\star$, where $X^\star$ is given by Proposition 3.9(iv). On the other hand, for the numerical stability of the MILP solver, $\overline{\boldsymbol{x}}$ should not be set too large.
- $\underline{\phi}$ specifies an initial lower bound of $\phi(f)$. In order to guarantee $\underline{\phi} \leq \phi(f)$, one finds c_0 and $\boldsymbol{y}_0$ such that $c_0 + \langle \boldsymbol{y}_0, \boldsymbol{g} \rangle \geq -f$ and set $\underline{\phi} = -c_0 - \pi(\boldsymbol{y}_0)$. The existence of such c_0 and $\boldsymbol{y}_0$ is guaranteed by Assumption 3.6(iii). Under Assumption 2.1, one can show that $\underline{\phi} \leq \phi(f)$ always holds. This is detailed in Theorem 3.16.
- ε is a positive number indicating the numerical accuracy of the algorithm. This is detailed in Theorem 3.16(ii).

- τ can be set as any positive number to provide a strict lower bound on $\phi(f)$.
- $\delta \in (0, 1]$ controls the number of additional constraints generated in each iteration of the algorithm in Line 13, each of which corresponds to a sub-optimal integer feasible solution found by the BnB algorithm before the optimizer (see Remark 3.13). When δ is set to 1, $X^{(r)}$ will be a singleton and contains only the $\boldsymbol{x}^\star$ that corresponds to an optimizer of problem (3.14). When δ is set close to 0, $X^{(r)}$ corresponds to most of the integer feasible solutions with $\boldsymbol{x}$ such that $c^{(r)} + s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x}) < 0$ and thus more constraints are included in the next LP problem in Line 9. δ can usually be set small, since we empirically observe that it often results in a reduction of the total number of iterations.

Remark 3.15. Algorithm 1 is inspired by the Conceptual Algorithm 11.4.1 in Goberna and López [37]. There are a few notable differences, as detailed below.

- In Line 2, the initial constraints involve auxiliary variables from the radial constraints generated by Algorithm 0.
- Line 4 introduces a lower bound on $\phi(f)$ to guarantee that the LP problem in Line 9 is always bounded.
- Line 13 uses sub-optimal integer feasible solutions of the MILP problem (3.14), allowing for more than one cuts to be generated in each iteration of Algorithm 1, which significantly speeds up the algorithm.
- In contrast to the Conceptual Algorithm 11.4.1 in Goberna and López [37] which returns an infeasible solution in general, Algorithm 1 returns a feasible solution due to Line 15. This will be proved below in Theorem 3.16.

Let us recall the definition of the feasible set S in (3.2) and the definition of the set of optimizers $S^\star$ in (3.3).

Theorem 3.16 (Properties of Algorithm 1)**.** *Let Assumption 3.6 hold. Assume that $\overline{\boldsymbol{x}}$ and $\underline{\phi}$ are specified as stated in Remark 3.14. Then, the following statements hold:*

(i) If Assumption 2.1 holds, then $\underline{\varphi}^{(r)}$ is non-decreasing in r. At any stage of Algorithm 1, $s^{(r)} \leq 0$ and $\underline{\varphi}^{(r)} \leq \phi(f) \leq \underline{\varphi}^{(r)} - s^{(r)}$.

(ii) If Assumption 2.1 holds, then Algorithm 1 terminates after finitely many iterations with an ε-optimal solution $(c^\star, \boldsymbol{y}^\star)$ of (2.4) and $\phi(f)^{\mathrm{LB}} \leq \phi(f) \leq \phi(f)^{\mathrm{UB}}$ with $\phi(f)^{\mathrm{UB}} - \phi(f)^{\mathrm{LB}} \leq \varepsilon$.

(iii) If Line 17 of Algorithm 1 is reached, then Assumption 2.1 is violated and problem (3.1) is unbounded.

Proof. See Appendix B.1.2. □

Remark 3.17. In Section 3.2, under the more restrictive assumption that $\Omega = \{\boldsymbol{x} \in \mathbb{R}^d : \mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}\}$ for some $\overline{\boldsymbol{x}} := (\overline{x}_1, \ldots, \overline{x}_d)^\mathsf{T} > \mathbf{0}$ (see Assumption 3.18), we show that Algorithm 1 also produces an ε-optimal solution to the right-hand side of (2.8), which corresponds to the most extreme pricing measure in the original model-free superhedging problem. This will be explained in detail in Corollary 3.25.

3.2. **CPWA payoff functions on bounded domains.** In the second setting, we adopt similar but more restrictive assumptions than in the first one.

Assumption 3.18 (Setting 2)**.** We assume the following:

(i) $\Omega = \{\boldsymbol{x} \in \mathbb{R}^d : \mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}\}$ for $\overline{\boldsymbol{x}} := (\overline{x}_1, \ldots, \overline{x}_d)^\mathsf{T} > \mathbf{0}$;
(ii) f and $(g_j)_{j=1:m}$ are CPWA functions on Ω;
(iii) $\phi(f) < \infty$ and $\phi(-f) < \infty$.

Remark 3.19. The ECP method (Algorithm 1) introduced in Section 3.1 still works under Assumption 3.18. The difference is that now the slack function $s_{\boldsymbol{y}}$ is always bounded from below on Ω, hence Line 1 and Line 2 are skipped in Algorithm 1. One potential drawback of Algorithm 1 is that the MILP problem (Line 12) needs to be solved till termination. This could be computationally inefficient when the size of the MILP problem is large. Another potential drawback of Algorithm 1 is that the number of constraints might become large, which makes solving the LP problem in Line 9 time-consuming.

Because of the drawbacks of Algorithm 1 discussed in Remark 3.19, we will introduce a version of the accelerated central cutting plane (ACCP) method based on Betrò [11], detailed in Algorithm 2. In Algorithm 2, we maintain and update a sequence of lower bounds $(\underline{\varphi}^{(r)})_{r\geq 0}$ of $\phi(f)$, a sequence of upper bounds $(\overline{\varphi}^{(r)})_{r\geq 0}$ of $\phi(f)$, and polytopes in $\mathbb{R}^{2m+1}$ that are denoted by $\sigma(\overline{c},\overline{\boldsymbol{y}},\underline{\varphi}^{(r)},\varphi^{(r)},X)$, which have the form

$$\begin{aligned}\sigma(\overline{c},\overline{\boldsymbol{y}},\underline{\varphi}^{(r)},\varphi^{(r)},X) := \Big\{(c,\boldsymbol{y}^+,\boldsymbol{y}^-) :&|c|\leq\overline{c},\mathbf{0}\leq\boldsymbol{y}^+\leq\overline{\boldsymbol{y}},\mathbf{0}\leq\boldsymbol{y}^-\leq\overline{\boldsymbol{y}},\\ &\underline{\varphi}^{(r)}\leq c+\langle\boldsymbol{y}^+,\overline{\boldsymbol{\pi}}\rangle-\langle\boldsymbol{y}^-,\underline{\boldsymbol{\pi}}\rangle\leq\varphi^{(r)},\\ &c+\langle\boldsymbol{y}^+-\boldsymbol{y}^-,\boldsymbol{g}(\boldsymbol{x})\rangle\geq f(\boldsymbol{x}),\ \forall\boldsymbol{x}\in X\Big\},\end{aligned} \tag{3.16}$$

where $\overline{c},\overline{\boldsymbol{y}}$ specify a bounding box, $\underline{\varphi}^{(r)}$ and $\overline{\varphi}^{(r)}$ specify the lower and upper bounds on $c+\pi(\boldsymbol{y}^+-\boldsymbol{y}^-)$, which are referred to as lower/upper objective cuts, and $X\subset\Omega$ specifies a collection of feasibility constraints, which are referred to as feasibility cuts. In Algorithm 2, $\varphi^{(r)}$ is between the lower bound $\underline{\varphi}^{(r)}$ and the upper bound $\overline{\varphi}^{(r)}$, and is used as a speculative upper objective cut. The idea of Algorithm 2 is that $(\underline{\varphi}^{(r)})_{r\geq 0}$ is a non-decreasing sequence of lower bounds that approaches $\phi(f)$ from below while $(\overline{\varphi}^{(r)})_{r\geq 0}$ is a non-increasing sequence of upper bounds that approaches $\phi(f)$ from above. These will be made clear later in Theorem 3.24. Algorithm 2 has various advantages over Algorithm 1. Most importantly, Algorithm 2 produces a sequence of feasible solutions and a sequence of non-increasing upper bounds approaching $\phi(f)$ from above (see Theorem 3.24). Even though in Algorithm 1, $(\underline{\varphi}^{(r)}-s^{(r)})_{r\geq 0}$ are upper bounds of $\phi(f)$ by Theorem 3.16(i), they are not non-increasing and it might occur that $|s^{(r)}|$ remains large before Algorithm 1 terminates. Moreover, the MILP problem in Line 20 of Algorithm 2 only needs to be solved approximately with a large error tolerance, and some linear constraints are removed in Line 34 to make solving the LP problem in Line 11 faster. We refer the reader to Betrò [11] and Section 11.4 of Goberna and López [37] for further discussions.

A crucial step of Algorithm 2 is to compute the Chebyshev center, that is, the center of the largest inscribed ball of the polytope $\sigma(\overline{c},\overline{\boldsymbol{y}},\underline{\varphi}^{(r)},\varphi^{(r)},X)$ in Line 11. It is well-known that the Chebyshev center of a polytope can be computed by solving an LP problem (see, for example, the Conceptual Algorithm 11.4.2 of Goberna and López [37]), which is given as a lemma below.

Lemma 3.20. *Let a polytope C be defined by $C := \{\boldsymbol{w} : \langle\boldsymbol{a}_j,\boldsymbol{w}\rangle\leq b_j, 1\leq j\leq J\}$ for $\boldsymbol{a}_j\in\mathbb{R}^n, b_j\in\mathbb{R}$, $j=1,\ldots,J$. Assume that C is non-empty. Let $(\rho^\star,\boldsymbol{w}^\star)$ be an optimal solution of the following LP problem:*

$$\begin{aligned}\textit{minimize}\quad & \rho\\ \textit{subject to}\quad & \langle\boldsymbol{a}_j,\boldsymbol{w}\rangle+\|\boldsymbol{a}_j\|\rho\leq b_j,\ \textit{for } j=1,\ldots,J,\\ & \rho\geq 0.\end{aligned} \tag{3.17}$$

Then $\boldsymbol{w}^\star$ is a Chebyshev center of C and $\rho^\star$ is the corresponding radius of the largest inscribed ball. If C is empty, then the LP problem (3.17) is infeasible.

Remark 3.21. Lemma 3.20 is not applicable to a P-polytope resulting from a projection (*e.g.* a system of linear inequalities containing auxiliary variables). Thus, Algorithm 2 is not applicable in Setting 1 (Assumption 3.6). Zhen and Den Hertog [65] made a detailed exposition of the various approaches to approximately compute the Chebyshev center of such a polytope. However, the computational cost of these approaches is too high to be considered for our problem.

Algorithm 2: Accelerated Central Cutting Plane (ACCP) Algorithm (under Assumption 3.18)

Input: $\overline{\boldsymbol{\pi}}$, $\underline{\boldsymbol{\pi}}$, $(g_j)_{j=1:m}$, f, $X^{(0)} \subset \Omega$, $\overline{\boldsymbol{x}}$, $\underline{\phi}$, $\overline{\phi}$, $c^{\star(0)}$, $\boldsymbol{y}^{\star(0)}$, $\overline{c} > 0$, $\overline{\boldsymbol{y}} > \mathbf{0}$, $\varepsilon > 0$, $\tau > \varepsilon$, $0 \leq \gamma < 1$, $0 < \zeta < 1$, $0 < \delta \leq 1$

Output: $\phi(f)^{\mathrm{UB}}$, $\phi(f)^{\mathrm{LB}}$, $c^\star$, $\boldsymbol{y}^\star$, $c^\dagger$, $\boldsymbol{y}^\dagger$, $X^\dagger$, X

1 $\underline{\varphi}^{(0)} \leftarrow \underline{\phi} - \tau$, $\overline{\varphi}^{(0)} \leftarrow \overline{\phi}$.
2 flag $\leftarrow$ false.
3 Mark all elements of $X^{(0)}$ as active and removable.
4 $r \leftarrow 0$.
5 **while** $\overline{\varphi}^{(r)} - \underline{\varphi}^{(r)} > \varepsilon$
6     $r \leftarrow r + 1$.
7     $\underline{\varphi}^{(r)} \leftarrow \underline{\varphi}^{(r-1)}$, $\overline{\varphi}^{(r)} \leftarrow \overline{\varphi}^{(r-1)}$, $\varphi^{(r)} \leftarrow (\underline{\varphi}^{(r)} + \overline{\varphi}^{(r)})/2$, $c^{\star(r)} \leftarrow c^{\star(r-1)}$, $\boldsymbol{y}^{\star(r)} \leftarrow \boldsymbol{y}^{\star(r-1)}$.
8   **if** *flag is true* **then**
9         $\varphi^{(r)} \leftarrow (\underline{\varphi}^{(r)} + \varphi^{(r)})/2$.
10     $X \leftarrow \bigcup_{0 \leq l \leq r-1} \{\boldsymbol{x} \in X^{(l)} : \boldsymbol{x} \text{ is marked active}\}$.
11   Compute the Chebyshev center of $\sigma(\overline{c}, \overline{\boldsymbol{y}}, \underline{\varphi}^{(r)}, \varphi^{(r)}, X)$ by solving the LP problem in (3.17).
12   **if** *the LP problem in Line 11 is infeasible* **then**
13         $\underline{\varphi}^{(r)} \leftarrow \min\{c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}} \rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}} \rangle : (c, \boldsymbol{y}^+, \boldsymbol{y}^-) \text{ satisfies } \sigma(\overline{c}, \overline{\boldsymbol{y}}, -\infty, \infty, X)\}$ (which is an LP problem). Let $(c^\dagger, \boldsymbol{y}^{+\dagger}, \boldsymbol{y}^{-\dagger})$ be a minimizer of the LP problem.
14         $\boldsymbol{y}^\dagger \leftarrow \boldsymbol{y}^{+\dagger} - \boldsymbol{y}^{-\dagger}$, $X^\dagger \leftarrow X$.
15   Mark all elements of $\bigcup_{1 \leq l \leq r-1} X^{(r)}$ as removable.
16         $\rho^{(r)} \leftarrow -1$, $X^{(r)} \leftarrow \emptyset$.
17   Skip to the next iteration.
18   Let $(c^{(r)}, \boldsymbol{y}^{+(r)}, \boldsymbol{y}^{-(r)})$ be the Chebyshev center and let $\rho^{(r)}$ be the radius of the largest inscribed ball of $\sigma(\overline{c}, \overline{\boldsymbol{y}}, \underline{\varphi}^{(r)}, \varphi^{(r)}, X)$.
19     $\boldsymbol{y}^{(r)} \leftarrow \boldsymbol{y}^{+(r)} - \boldsymbol{y}^{-(r)}$.
20   Formulate the global minimization problem: minimize $c^{(r)} + s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x})$ subject to $\mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}$ into the MILP problem in (3.14). Solve it with relative gap tolerance ζ. Let $\overline{s}^{(r)}$ be its approximate optimal value. Let $\underline{s}^{(r)}$ be its lower bound at termination.
21     $X^{(r)} \leftarrow \Big\{\boldsymbol{x} : (\boldsymbol{x}, (\lambda_k), (\zeta_k), (\delta_{k,i}), (\iota_{k,i})) \text{ is an integer feasible solution found by the BnB algorithm while solving (3.14) such that } c^{(r)} + s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x}) \leq \delta\overline{s}^{(r)}\Big\}$. Mark all elements of $X^{(r)}$ as active and removable.
22   **if** $c^{(r)} + \pi(\boldsymbol{y}^{(r)}) - \underline{s}^{(r)} < \overline{\varphi}^{(r)} - \varepsilon$ **then**
23         $\overline{\varphi}^{(r)} \leftarrow c^{(r)} + \pi(\boldsymbol{y}^{(r)}) - \underline{s}^{(r)}$, $c^{\star(r)} \leftarrow c^{(r)} - \underline{s}^{(r)}$, $\boldsymbol{y}^{\star(r)} \leftarrow \boldsymbol{y}^{(r)}$.
24   **if** $\underline{s}^{(r)} \geq 0$ **then**
25   Mark all elements of $\bigcup_{1 \leq l \leq r} X^{(l)}$ as removable.
26   Skip to the next iteration.
27   **if** *flag is true* **then**
28   flag $\leftarrow$ false.
29   Mark all elements of $X^{(r)}$ as non-removable.
30   Skip to the next iteration.
31   flag $\leftarrow$ true.

(continues on the next page)

Algorithm 2: Accelerated Central Cutting Plane (ACCP) Algorithm (under Assumption 3.18)

(continued)

31 **for** *each* $0 \leq l \leq r$ *such that* $\rho^{(r)} < \gamma\rho^{(l)}$ **do**

32 **for** *each* $\boldsymbol{x} \in X^{(l)}$ *marked as removable* **do**

33 **if** *the constraint in (3.17) corresponding to* $\boldsymbol{x}$ *is satisfied strictly by* $(c^{(r)}, \boldsymbol{y}^{+(r)}, \boldsymbol{y}^{-(r)}, \rho^{(r)})$ *when solving the LP problem in Line 11* **then**

34 Set $\boldsymbol{x}$ as inactive.

35 $\phi(f)^{\mathrm{UB}} \leftarrow \overline{\varphi}^{(r)}$, $\phi(f)^{\mathrm{LB}} \leftarrow \underline{\varphi}^{(r)}$, $c^\star \leftarrow c^{\star(r)}$, $\boldsymbol{y}^\star \leftarrow \boldsymbol{y}^{\star(r)}$, $X \leftarrow \bigcup_{0 \leq l \leq r} \{\boldsymbol{x} \in X^{(l)} : \boldsymbol{x} \text{ is marked active}\}$.

36 **if** $\phi(f)^{\mathrm{UB}} < \underline{\phi}$ **then**

37 **return** The problem (3.1) is unbounded.

38 **else**

39 **return** $\phi(f)^{\mathrm{UB}}$, $\phi(f)^{\mathrm{LB}}$, $c^\star$, $\boldsymbol{y}^\star$, $c^\dagger$, $\boldsymbol{y}^\dagger$, $X^\dagger$, X.

In the following, Remark 3.22 explains the inputs to Algorithm 2, Remark 3.23 explains some of its unique features, and Theorem 3.24 shows its validity.

Remark 3.22 (Inputs to Algorithm 2)**.** The following list explains the various inputs to Algorithm 2 and how to set them.

- $\overline{\boldsymbol{\pi}}$, $\underline{\boldsymbol{\pi}}$, $(g_j)_{j=1:m}$, f are given by problem (3.1).
- $X^{(0)}$ is the same as in Algorithm 1 and Remark 3.14.
- $\overline{\boldsymbol{x}}$ is specified in Assumption 3.18.
- $\underline{\phi}$ specifies an initial lower bound of $\phi(f)$ and is obtained in the same way as in Remark 3.14.
- $(c^{\star(0)}, \boldsymbol{y}^{\star(0)})$ specifies an initial feasible point of (2.4), *i.e.* $\binom{c^{\star(0)}}{\boldsymbol{y}^{\star(0)}} \in S$, and $\overline{\phi} = c^{\star(0)} + \pi(\boldsymbol{y}^{\star(0)})$ is an initial upper bound of $\phi(f)$.
- $\overline{c}$, $\overline{\boldsymbol{y}}$ specify a bounding box on $(c, \boldsymbol{y}^+, \boldsymbol{y}^-)$ as in (3.16). To guarantee the validity of Algorithm 2, $\overline{c}$ and $\overline{\boldsymbol{y}}$ need to be specified such that $\{\binom{c}{\boldsymbol{y}} : |c| \leq \overline{c}, -\overline{\boldsymbol{y}} \leq \boldsymbol{y} \leq \overline{\boldsymbol{y}}\} \cap S^\star \neq \emptyset$.
- ε is a positive number specifying the numerical accuracy of the algorithm. This is detailed in Theorem 3.24(ii) and (iii).
- τ can be set as any number greater than ε to provide a strict lower bound on $\phi(f)$.
- $\gamma \in [0, 1)$ controls the removal of active feasibility cuts in Line 34. When γ is set to 0, then Line 34 is never reached and all feasibility cuts are retained. If γ is set close to 1, then more feasibility cuts are removed, making solving the LP problem in Line 11 and Line 13 faster.
- $\zeta \in (0, 1)$ controls the termination of the BnB algorithm, hence also the solution quality of the MILP problem in Line 20. Suppose that the true optimal value of the MILP problem is $s^{(r)}$. As in Remark 3.13, in the BnB algorithm, $\overline{s}^{(r)}$ is the approximate optimal value, and $\underline{s}^{(r)}$ is the best objective bound at termination. By the termination condition in (3.15), $\frac{\overline{s}^{(r)} - \underline{s}^{(r)}}{|\overline{s}^{(r)}|} \leq \zeta$. If $\overline{s}^{(r)} \geq 0$, then $\overline{s}^{(r)} \geq s^{(r)} \geq \underline{s}^{(r)} \geq 0$. If $\overline{s}^{(r)} < 0$, then $(1+\zeta)\overline{s}^{(r)} \leq \underline{s}^{(r)} \leq s^{(r)} \leq \overline{s}^{(r)} < 0$, hence $s^{(r)} \leq \overline{s}^{(r)} \leq \frac{1}{1+\zeta} s^{(r)}$.
- δ is the same as in Algorithm 1 and Remark 3.14.

Remark 3.23. Algorithm 2 is partially based on the accelerated central cutting plane algorithm in Section 2 of Betrò [11], with the following differences:

- Instead of starting with $\underline{\varphi}^{(0)} = -\infty$, Algorithm 2 starts with a given lower bound $\underline{\varphi}^{(0)} > -\infty$ similar to Algorithm 1.

- In Algorithm 2, whenever the current polytope $\sigma(\overline{c}, \overline{\boldsymbol{y}}, \underline{\varphi}^{(r)}, \varphi^{(r)}, X)$ is found to be empty, instead of setting $\underline{\varphi}^{(r)} \leftarrow \varphi^{(r)}$, an LP problem is solved in Line 13 to update the lower bound $\underline{\varphi}^{(r)}$.
- Similar to Algorithm 1, in Line 21 of Algorithm 2, sub-optimal integer feasible solutions found by the MILP solver are used to generate additional feasibility cuts to speed up the algorithm.
- In Line 23 of Algorithm 2, a feasible point of the LSIP problem is generated from a possibly infeasible one by setting $c^{\star(r)} \leftarrow c^{(r)} - \underline{s}^{(r)}$. This can result in an objective cut being made even when the current Chebyshev center is infeasible.

Theorem 3.24 (Properties of Algorithm 2)**.** *Let Assumption 3.18 hold. Assume that $\underline{\phi}$, $\overline{\phi}$, $c^{\star(0)}$, $\boldsymbol{y}^{\star(0)}$, $\overline{c}$, $\overline{\boldsymbol{y}}$ are specified as stated in Remark 3.22. Then, the following statements hold:*

(i) If Assumption 2.1 holds, then $\underline{\varphi}^{(r)}$ is non-decreasing in r, $\overline{\varphi}^{(r)}$ is non-increasing in r. Moreover, at any stage of Algorithm 2, $\underline{\varphi}^{(r)} \le \phi(f) \le \overline{\varphi}^{(r)}$, and $c^{\star(r)} + \langle \boldsymbol{y}^{\star(r)}, \boldsymbol{g}\rangle \ge f$ holds.

(ii) If Assumption 2.1 holds, then Algorithm 2 terminates after finitely many iterations with a feasible and ε-optimal solution $(c^\star, \boldsymbol{y}^\star)$ and $\phi(f)^{\mathrm{LB}} \le \phi(f) \le \phi(f)^{\mathrm{UB}}$ with $\phi(f)^{\mathrm{UB}} - \phi(f)^{\mathrm{LB}} \le \varepsilon$.

(iii) If Assumption 2.1 holds, then $c^\dagger$, $\boldsymbol{y}^\dagger$ and $X^\dagger$ are defined when Algorithm 2 terminates. If $|c^\dagger| < \overline{c}$ and $-\overline{\boldsymbol{y}} < \boldsymbol{y}^\dagger < \overline{\boldsymbol{y}}$, then the following LP problem with decision variables $(\mu_{\boldsymbol{x}})_{\boldsymbol{x}\in X^\dagger}$:

$$\begin{aligned} \text{maximize} \quad & \sum_{\boldsymbol{x}\in X^\dagger} \mu_{\boldsymbol{x}} f(\boldsymbol{x}) \\ \text{subject to} \quad & \sum_{\boldsymbol{x}\in X^\dagger} \mu_{\boldsymbol{x}} = 1, \\ & \underline{\boldsymbol{\pi}} \le \sum_{\boldsymbol{x}\in X^\dagger} \mu_{\boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}) \le \overline{\boldsymbol{\pi}}, \\ & \mu_{\boldsymbol{x}} \ge 0, \ \forall \boldsymbol{x} \in X^\dagger \end{aligned} \tag{3.18}$$

has an optimal solution $(\mu^\star_{\boldsymbol{x}})_{\boldsymbol{x}\in X^\dagger}$. Define the finitely supported measure $\mu^\star$ by $\mu^\star := \sum_{\boldsymbol{x}\in X^\dagger} \mu^\star_{\boldsymbol{x}} \delta_{\boldsymbol{x}}$. Then $\mu^\star$ is ε-optimal for the right-hand side of (2.8), which corresponds to the most extreme pricing measure in the original model-free superhedging problem.

(iv) If Line 37 of Algorithm 2 is reached, then Assumption 2.1 is violated and problem (3.1) is unbounded.

Proof. See Appendix B.2. □

Theorem 3.24(iii) explicitly provides a pricing measure which is an ε-optimal solution to the original model-free superhedging problem. The ECP method (Algorithm 1) also has this property in Setting 2 (Assumption 3.18), which is detailed in the next corollary.

Corollary 3.25. *Under Assumption 2.1 and Assumption 3.18, Algorithm 1 (with Line 1 and Line 2 removed) terminates after finitely many iterations, and the following LP problem with decision variables $(\mu_{\boldsymbol{x}})_{\boldsymbol{x}\in X}$:*

$$\begin{aligned} \text{maximize} \quad & \sum_{\boldsymbol{x}\in X} \mu_{\boldsymbol{x}} f(\boldsymbol{x}) \\ \text{subject to} \quad & \sum_{\boldsymbol{x}\in X} \mu_{\boldsymbol{x}} = 1, \\ & \underline{\boldsymbol{\pi}} \le \sum_{\boldsymbol{x}\in X} \mu_{\boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}) \le \overline{\boldsymbol{\pi}}, \\ & \mu_{\boldsymbol{x}} \ge 0, \ \forall \boldsymbol{x} \in X \end{aligned} \tag{3.19}$$

has an optimal solution $(\mu^\star_{\boldsymbol{x}})_{\boldsymbol{x}\in X}$. *Define the finitely supported measure* $\mu^\star$ *by* $\mu^\star := \sum_{\boldsymbol{x}\in X} \mu^\star_{\boldsymbol{x}}\delta_{\boldsymbol{x}}$. *Then* $\mu^\star$ *is* ε*-optimal for the right-hand side of (2.8).*

Proof. See Appendix B.2. □

3.3. **Comparison with integral-type penalization methods.** A major limitation of the discretization-based approaches for solving LSIP problems (*e.g.* Algorithm 1 and Algorithm 2) is the difficulty in solving the global non-convex optimization problems involved in these approaches. Specifically, in our problem, solving the MILP problem in Line 12 of Algorithm 1 and in Line 20 of Algorithm 2 can be computationally costly. There is an alternative approach based on integral-type penalization for solving semi-infinite optimization problems, see *e.g.* Auslender, Goberna, and López [3], Borwein and Lewis [12, 13], Lin, Fang, and Wu [47]. This approach transforms the original problem into an unconstrained convex optimization problem with an integral-type penalty term and circumvents the need to solve the difficult global non-convex optimization problem. Recently, a similar approach has been adopted for solving optimal transport, martingale optimal transport and related problems, see *e.g.* Aquino and Bernard [2], Eckstein and Kupper [32], Eckstein, Kupper, and Pohl [33], Eckstein et al. [34].

More specifically, in the integral-type penalization approach, one solves the following penalized version of the LSIP problem (3.1):

$$\phi_\gamma(f) := \inf_{c\in\mathbb{R},\, \boldsymbol{y}\in\mathbb{R}^m} \left\{ c + \pi(\boldsymbol{y}) + \gamma \int_\Omega \beta\left(\frac{1}{\gamma}\left[f(\omega) - c - \langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle\right]\right)\theta(\mathrm{d}\omega)\right\}, \tag{3.20}$$

where $\beta : \mathbb{R} \mapsto \mathbb{R}_+$ is a convex, non-decreasing, and continuously differentiable function which satisfies $\lim_{t\to\infty} \frac{\beta(t)}{t} \to \infty$, $\gamma > 0$, and θ is a positive finite Borel measure on Ω. Under Assumption 3.18, and the additional assumption that θ is a probability measure with positive density with respect to the Lebesgue measure on Ω, one is able to show the following duality by a similar argument as in the proof of Theorem 2.4(iii), and by using Theorem 4.2 of Borwein and Lewis [14] and Theorem 2 of Rockafellar [59]:

$$\phi_\gamma(f) = \sup_{\rho\in C_+,\, \rho\mathrm{d}\theta\in\mathcal{Q}} \left\{ \int_\Omega f(\omega)\rho(\omega)\theta(\mathrm{d}\omega) - \gamma\int_\Omega \beta^*\left(\rho(\omega)\right)\theta(\mathrm{d}\omega)\right\}, \tag{3.21}$$

where $C_+ := \{\rho \in \mathcal{L}^1(\Omega,\theta) : \rho \geq 0\ \theta\text{-a.e.}, \int_\Omega (|f(\omega)| + \sum_{j=1}^m |g_j(\omega)|)\rho(\omega)\theta(\mathrm{d}\omega) < \infty\}$, and $\beta^*(\lambda) := \sup_{t\in\mathbb{R}}\{\lambda t - \beta(t)\}$ is the convex conjugate of β. Similar results can be found in Borwein and Lewis [13]. One may view (3.21) as a regularized version of the model-free superhedging problem $\sup_{\mu\in\mathcal{Q}} \int_\Omega f \,\mathrm{d}\mu$ (see *e.g.* Lin et al. [47]). Compared to Algorithm 1 and Algorithm 2, the integral-type penalization approach does not produce feasible and ε-optimal solutions to (3.1) and $\sup_{\mu\in\mathcal{Q}} \int_\Omega f\,\mathrm{d}\mu$ like what we have shown in Theorem 3.16, Theorem 3.24 and Corollary 3.25 due to the relaxation of the constraint $c + \langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle \geq f(\omega)$ for all $\omega \in \Omega$.

A challenge in this approach is the difficulty in solving (3.20) numerically. The penalty term $\int_\Omega \beta\left(\frac{1}{\gamma}\left[f(\omega) - c - \langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle\right]\right)\theta(\mathrm{d}\omega)$ involves an integral which is analytically intractable and computationally very hard to evaluate. One possible way to tackle this challenge is to perform Monte Carlo integration and further transform the problem into an unconstrained convex optimization problem with stochastic objective. Subsequently, the stochastic (sub)gradient descent (SGD) algorithm can be used to solve the problem. This is similar to the approach adopted in Aquino and Bernard [2], Eckstein and Kupper [32], Eckstein et al. [33, 34].

Remark 3.26. The difference between our problem and Aquino and Bernard [2], Eckstein and Kupper [32], Eckstein et al. [33, 34] is that these studies solve multi-marginal optimal transport, martingale optimal transport and related problems where the marginals are known and fixed, hence the decision space of the optimization problem contains sets of continuous functions which are parametrized by artificial neural networks. As a result, the stochastic objective functions in Aquino and Bernard [2], Eckstein and Kupper [32], Eckstein et al. [33, 34] are non-convex.

We have empirically tested this *penalization plus SGD* approach in our problem, and found that it is highly sensitive to the initialization and the hyper-parameter settings of the SGD algorithm.

Moreover, when the dimension of Ω is high, for instance, when $d \geq 10$, we have experienced the following dilemma: when γ is chosen to be very close to zero, the Monte Carlo approximation of $\int_\Omega \beta\left(\frac{1}{\gamma}[f(\omega) - c - \langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle]\right)\theta(\mathrm{d}\omega)$ has extremely large variance and the SGD algorithm does not converge; when γ is chosen to be large enough to permit convergence of the SGD algorithm, $\phi_\gamma(f)$ is often far smaller than $\phi(f)$.

4. Numerical experiments and results

In the previous section, we introduced two different settings and two numerical algorithms for computing model-free bounds for multi-asset option prices in the presence of option-implied information. In the first setting, we assumed that the prices of the assets underlying the derivatives are non-negative, and developed the exterior cutting plane (ECP) method for the computations. In the second setting, we assumed that prices are non-negative and upper bounded, and developed the accelerated central cutting plane (ACCP) method for the computations, while the ECP method still applies. In this section, we perform experiments using market prices of assets and derivatives that are synthetically generated, in order (i) to demonstrate the performance of the proposed approaches under these settings, and (ii) to quantify the effect of the additional information from traded multi-asset options on the width of the no-arbitrage gap, *i.e.* the difference between the upper and lower model-free bounds.

Details of implementation. All algorithms and experiments are implemented using the MATLAB language. We adopt the MATLAB interface of Gurobi to solve linear programs and mixed-integer linear programs within the algorithms. The MATLAB code used in this work is available on GitHub[2].

The list below briefly discusses some numerical considerations in the implementation of the algorithms. The details of these considerations can be found in the MATLAB code.

- All LP problems in Algorithm 1 and Algorithm 2 are solved using the interior point algorithm for better numerical accuracy. Even though the dual simplex algorithm is usually more efficient, we observed empirically that it is numerically less accurate compared to the interior point algorithm.
- The efficiency and the numerical stability of the LP solver are sensitive to the numerical condition of the coefficients of the model, which is associated with the ratio between the largest and smallest coefficients in absolute values. Thus, we round the inputs of the model when constructing coefficient matrices of LP problems. Concretely, in Line 13 of Algorithm 1 and in Line 21 of Algorithm 2, each component of a solution $\boldsymbol{x}$ is rounded to a multiple of 10^{-4} to control the numerical condition of the coefficient matrix in the LP problem.
- When formulating the MILP problem in (3.14), we discard the terms in (3.13) in which $\boldsymbol{a}_{k,i}$ and $b_{k,i}$ are very close to zero (*i.e.* $\leq 10^{-6}$) to reduce the size of the resulting MILP problem.

Derivatives and models. In the subsequent numerical experiments, we consider financial derivatives with the following CPWA payoff functions, whose representations are discussed in Example 3.2.

(i) Trading in the i-th asset: $g(\boldsymbol{x}) = x_i$.
(ii) Vanilla call option on the i-th asset with strike $\kappa > 0$: $g(\boldsymbol{x}) = (x_i - \kappa)^+$.
(iii) Basket call option with weights $\boldsymbol{w} \in \mathbb{R}^d_+$ and strike $\kappa > 0$: $g(\boldsymbol{x}) = \left(\sum_i w_i x_i - \kappa\right)^+$.
(iv) Spread call option with weights $\boldsymbol{w} \in \mathbb{R}^d \setminus \mathbb{R}^d_+$ (*e.g.* $\boldsymbol{w} = \boldsymbol{e}_i - \boldsymbol{e}_j$) and strike $\kappa \in \mathbb{R}$, *e.g.*: $g(\boldsymbol{x}) = (x_i - x_j - \kappa)^+$.
(v) Call-on-max (rainbow) option on assets $i_1, \ldots, i_l$ with strike $\kappa \geq 0$:
$$g(\boldsymbol{x}) = \left(x_{i_1} \vee x_{i_2} \vee \ldots \vee x_{i_l} - \kappa\right)^+.$$
(vi) Call-on-min (rainbow) option of assets $i_1, \ldots, i_l$ with strike $\kappa \geq 0$:
$$g(\boldsymbol{x}) = \left(x_{i_1} \wedge x_{i_2} \wedge \ldots \wedge x_{i_l} - \kappa\right)^+.$$

[2]https://github.com/qikunxiang/ModelFreePriceBounds

(vii) Best-of-calls option of assets $i_1, \ldots, i_l$ with different strikes $\kappa_1, \ldots, \kappa_l \geq 0$:

$$g(\boldsymbol{x}) = (x_{i_1} - \kappa_1)^+ \vee (x_{i_2} - \kappa_2)^+ \vee \cdots \vee (x_{i_l} - \kappa_l)^+.$$

Moreover, in the numerical experiments, we consider market models of the following kind:

- The marginal distribution of the price of an asset at terminal time is log-normal with probability density function
$$p(x; \mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log x - \mu)^2}{2\sigma^2}\right).$$
This distribution involves two parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$.
- The dependence structure among the marginals of the d assets at terminal time is a t-copula with correlation matrix $\mathbf{C}$ and ν degrees of freedom.

Therefore, the complete characterization of a market model requires specifying the parameters of each of the marginal distributions, as well as the correlation coefficients and the degrees of freedom of the t-copula model.

Given this market model, the prices of the single-asset derivatives listed above can be computed in closed-form by taking the discounted expectations of the corresponding payoff functions (w.r.t. a pricing measure). We have assumed that the interest rate is equal to zero, for the sake of simplicity. For the multi-asset derivatives listed above, there is no closed-form formula to compute their prices, and we instead approximate them via Monte Carlo simulation by randomly generating 100,000 independent samples from the copula model and subsequently using these samples to approximate the expectations of the payoff functions. Since we focus on the case of an incomplete market with the presence of bid-ask spread, we also need to simulate the bid and ask prices of the derivatives. In order to do so, we specify multiple market models with different parameters. Subsequently, we compute the prices of the derivatives under each of the market models. The minimum (resp. maximum) price of a derivative among its prices under all models is taken as the bid (resp. ask) price of the derivative. Finally, we round the bid prices down to the nearest multiples of 0.001 and round the ask prices up to the nearest multiples of 0.001.

Notice that under the market models described above, the joint distribution of the asset prices corresponds to a probability measure $\widehat{\mu}$ on $\mathbb{R}^d_+$ with strictly positive density with respect to the Lebesgue measure. Moreover, the way the prices of derivatives are generated guarantees that $\widehat{\mu} \in \mathcal{Q}$. Therefore, by Proposition 2.9, Assumption 2.1 holds.

Remark 4.1. Under Setting 2, the way we generate the prices of derivatives is not completely correct, since the support of the joint distribution of the asset prices $\widehat{\mu}$ is in fact $\mathbb{R}^d_+$, which is larger than Ω. However, we can choose Ω such that $\mu(\mathbb{R}^d_+ \setminus \Omega)$ is negligible, and still guarantee that the generated derivative prices satisfy $\widehat{\mu} \in \mathcal{Q}$. In this case, Assumption 2.1 still holds by Proposition 2.9. Let us point out that this setting is neither unrealistic nor very restrictive, since asset and derivative prices are always finite.

4.1. **Experiment 1.** In this experiment, we consider a financial market with 5 assets ($d = 5$) and a fixed model, *i.e.* there is (almost) no bid-ask spread (see also Remark 4.2). We consider Setting 2 introduced in Section 3.2, and let $\Omega = [0, 100]^5$. Our goal is to compute the model-free lower and upper price bounds for a call-on-max option with payoff function:

$$f(\boldsymbol{x}) = \big(x_2 \vee x_3 \vee x_4 - \kappa\big)^+,$$

where the strike price κ ranges from 0 to 10 with an increment of 0.2. The purpose of this experiment is to demonstrate that the model-free price bounds can be improved, *i.e.* the no-arbitrage gap shrinks, by taking the market prices of increasingly more financial derivatives into account.

We assume that a total of 439 financial derivatives are traded in the market ($m = 439$). These include:

- The 5 assets x_1, x_2, x_3, x_4, x_5.
- Vanilla call options on the 5 assets with strikes $1, 2, \ldots, 10$.
- Basket call options with the following weights and strikes $1, 2, \ldots, 10$:

- $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})^\top$,
- $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0)^\top$, $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4})^\top$, $(\frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4}, \frac{1}{4})^\top$, $(\frac{1}{4}, 0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^\top$, $(0, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^\top$,
- $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, 0)^\top$, $(0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)^\top$, $(0, 0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})^\top$,
- $(0, \frac{1}{2}, \frac{1}{2}, 0, 0)^\top$, $(0, \frac{1}{2}, 0, \frac{1}{2}, 0)^\top$, $(0, 0, \frac{1}{2}, \frac{1}{2}, 0)^\top$.

- Spread call options with the following weights and strikes $-5, -4, \ldots, 0, 1, \ldots, 5$: $\boldsymbol{e}_1 - \boldsymbol{e}_2$, $\boldsymbol{e}_1 - \boldsymbol{e}_3$, $\boldsymbol{e}_1 - \boldsymbol{e}_4$, $\boldsymbol{e}_2 - \boldsymbol{e}_3$, $\boldsymbol{e}_2 - \boldsymbol{e}_4$, $\boldsymbol{e}_2 - \boldsymbol{e}_5$, $\boldsymbol{e}_3 - \boldsymbol{e}_4$, $\boldsymbol{e}_3 - \boldsymbol{e}_5$, $\boldsymbol{e}_4 - \boldsymbol{e}_5$, $\boldsymbol{e}_2 - \boldsymbol{e}_1$, $\boldsymbol{e}_3 - \boldsymbol{e}_1$, $\boldsymbol{e}_4 - \boldsymbol{e}_1$, $\boldsymbol{e}_3 - \boldsymbol{e}_2$, $\boldsymbol{e}_4 - \boldsymbol{e}_2$, $\boldsymbol{e}_5 - \boldsymbol{e}_2$, $\boldsymbol{e}_4 - \boldsymbol{e}_3$, $\boldsymbol{e}_5 - \boldsymbol{e}_3$, $\boldsymbol{e}_5 - \boldsymbol{e}_4$.
- Call-on-max (rainbow) options with the following 6 groups of assets and strikes $0, 1, \ldots, 10$: $\{x_1, x_2, x_3, x_4, x_5\}$, $\{x_1, x_2, x_3, x_4\}$, $\{x_2, x_3, x_4, x_5\}$, $\{x_2, x_3\}$, $\{x_2, x_4\}$, $\{x_3, x_4\}$.

The market model we are considering is specified as follows:

- The marginal distributions of the asset prices at terminal time are log-normal with different parameters.
- The dependence model is a t-copula with correlation matrix which has a 2-factor model structure:

$$\mathbf{C} = \mathbf{\Lambda}\mathbf{D}\mathbf{\Lambda}^\top + \mathbf{\Psi},$$

  where $\mathbf{\Lambda} \in \mathbb{R}^{d\times 2}$, $\mathbf{D} \in \mathbb{R}^{2\times 2}$ is a diagonal matrix, and $\mathbf{\Psi} \in \mathbb{R}^{d\times d}$ is a diagonal matrix. The degree of freedom parameter ν of the t-copula is set to 3.

After computing and rounding the prices of the assets and derivatives, we add 0.005 to and subtract 0.005 from these prices in order to get the ask and the bid prices. We also compute the price of the call-on-max option we are pricing with different strikes via Monte Carlo to serve as the ground truth.

Remark 4.2. Even though it is possible to modify the proposed algorithms slightly to handle the case without the bid-ask spread, we have empirically observed that doing so in this experiment will result in either numerical issues or finding violation of Assumption 2.1, *i.e.* reaching Line 17 of Algorithm 1 or Line 37 of Algorithm 2. The reason is that the Monte Carlo errors and numerical errors during the computation of the derivative prices make them violate or almost violate Assumption 2.1. Therefore, we add the ± 0.005 margin to the computed prices to account for such errors.

We consider 4 different cases, where we use certain subsets of the traded derivatives to compute the model-free lower and upper bounds:

- Case 1 (denoted as *V*): we use only vanilla options;
- Case 2 (denoted as *V+B*): we use vanilla and basket options;
- Case 3 (denoted as *V+B+S*): we use vanilla, basket, and spread options;
- Case 4 (denoted as *V+B+S+R*): we use vanilla, basket, spread and call-on-max (rainbow) options .

In order to compute the lower and upper bounds of the call-on-max option with payoff function f, we use the ECP method (Algorithm 1 with Line 1 and Line 2 removed) and the ACCP method (Algorithm 2). They approximately compute the model-free upper bound $\phi(f)$. The model-free lower bound is given by $-\phi(-f)$, which can also be approximately computed by replacing f with $-f$.

Let us now introduce the inputs to Algorithm 1 and Algorithm 2 in this experiment. In Algorithm 1, we set $\varepsilon = 0.001, \tau = 0.1, \delta = 0.3$. We set $\underline{\phi}$ according to Remark 3.14. When computing $\phi(f)$, we set $\underline{\phi} = 0$, which is given by $0 \geq -f$. When computing $\phi(-f)$, we set $\underline{\phi} = -\overline{\pi}_2 - \overline{\pi}_3 - \overline{\pi}_4$, which is given by $x_2 + x_3 + x_4 \geq f$. In Algorithm 2, we set $\varepsilon = 0.001, \tau = 0.1, \gamma = 0.1, \zeta = 0.01, \delta = 0.3, \overline{c} = 100, \overline{\boldsymbol{y}} = 100 \cdot \mathbf{1}$. $\underline{\phi}$ is set to be the same as in Algorithm 1 and $\overline{\phi}$ is set according to Remark 3.22. When computing $\phi(f)$, we set $\overline{\phi} = \overline{\pi}_2 + \overline{\pi}_3 + \overline{\pi}_4$ and $c^{\star(0)} + \langle \boldsymbol{y}^{\star(0)}, \boldsymbol{g}\rangle = x_2 + x_3 + x_4 \geq f$. When computing $\phi(-f)$, we set $\underline{\phi} = 0$ and $c^{\star(0)} + \langle \boldsymbol{y}^{\star(0)}, \boldsymbol{g}\rangle = 0 \geq -f$.

Figure 4.1 shows the computed lower and upper price bounds of the call-on-max option with different strikes, along with the prices simulated via Monte Carlo from the same market model that generated the prices of traded derivatives. The simulated prices represent the ground-truth in this experiment. The model-free lower and upper price bounds are computed by Algorithm 1 and Algorithm 2 and correspond to the outputs $-\phi(-f)^{\mathrm{UB}}$ and $\phi(f)^{\mathrm{UB}}$.

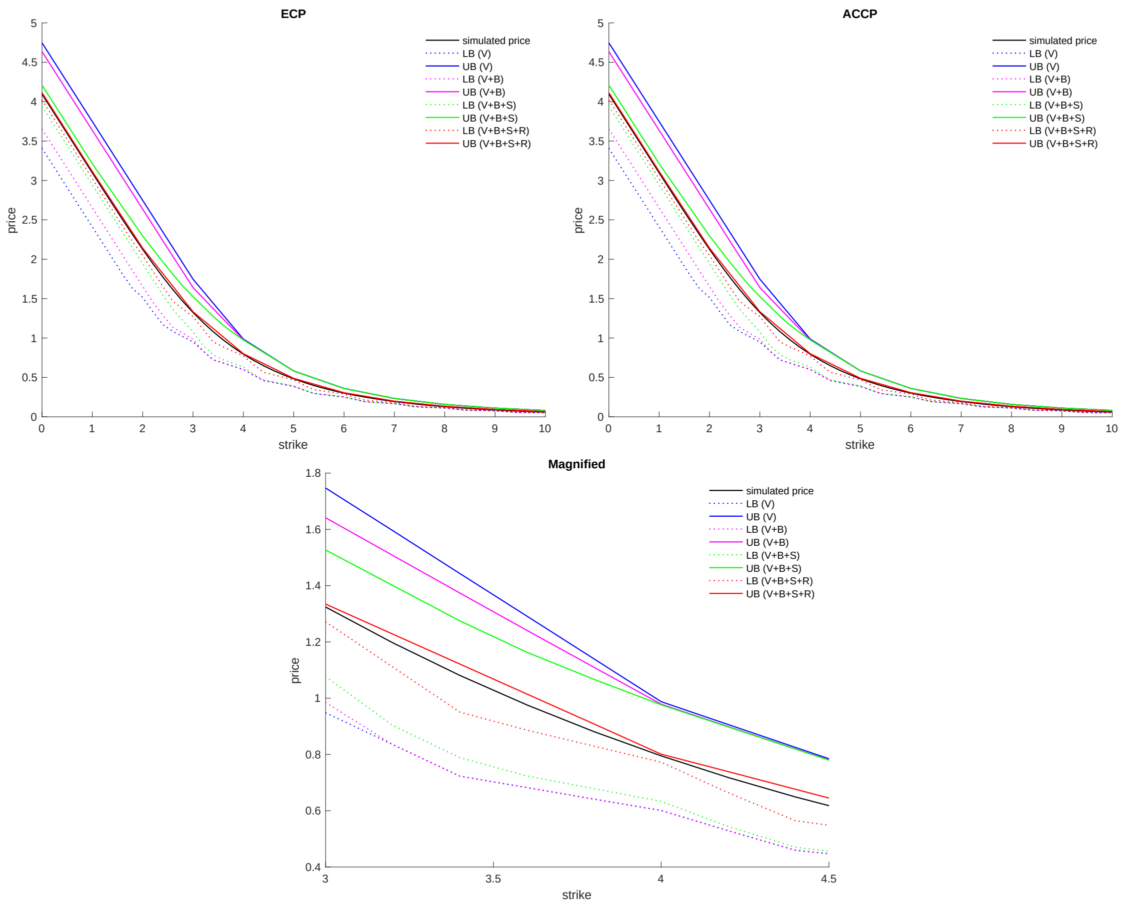

FIGURE 4.1. Experiment 1 – Model-free lower and upper price bounds of call-on-max options with strikes between 0 and 10 under 4 cases. The bottom panel shows a magnified version of a part of the top-right panel.

One can observe, that the price bounds resulted in the 4 cases are distinct, and that the gap between the lower and upper bounds shrinks when the prices of more traded derivatives are used. This tells us that observing the market prices of more traded derivatives substantially restricts the class of possible pricing measures $\mathcal{Q}$ and reduces the no-arbitrage gap between the bounds. When considering the corresponding dual optimization problem (2.4), this can be equivalently interpreted as having the information about more traded derivatives provides more ways to sub-replicate and super-replicate the given payoff function and thus makes the gap between the sub-replication price and the super-replication price smaller. In the Case *V+B+S+R* with all traded derivatives taken into consideration, the gap between the lower and upper bounds shrinks to almost zero. The reason is that the traded call-on-max options provide more information to determine the price of the target derivative, since they are similar in structure to the target derivative. This becomes concrete when one considers the dual optimization problem (2.4), where these call-on-max options offer direct ways to sub-replicate and super-replicate the target payoff, *e.g.* $\big(x_2 \vee x_3 - \kappa\big)^+ \le \big(x_2 \vee x_3 \vee x_4 - \kappa\big)^+ \le \big(x_1 \vee x_2 \vee x_3 \vee x_4 - \kappa\big)^+$. One may also notice that the price bounds computed by the two algorithms are almost identical. This is indeed the case since we have checked that all of the absolute differences between the bounds computed by the two algorithms are below $\varepsilon = 0.001$. This is a consequence of Theorem 3.16(ii) and Theorem 3.24(ii) and confirms the correctness of the computed price bounds.

4.2. **Experiment 2.** In this experiment, we keep the settings and the goal the same as in Experiment 1, while adding bid-ask spread to the prices of the traded derivatives. In order to do so, we introduce a second group of marginal distributions for the asset prices at terminal time that are different from the ones introduced in Section 4.1. Moreover, we introduce a second dependence model,

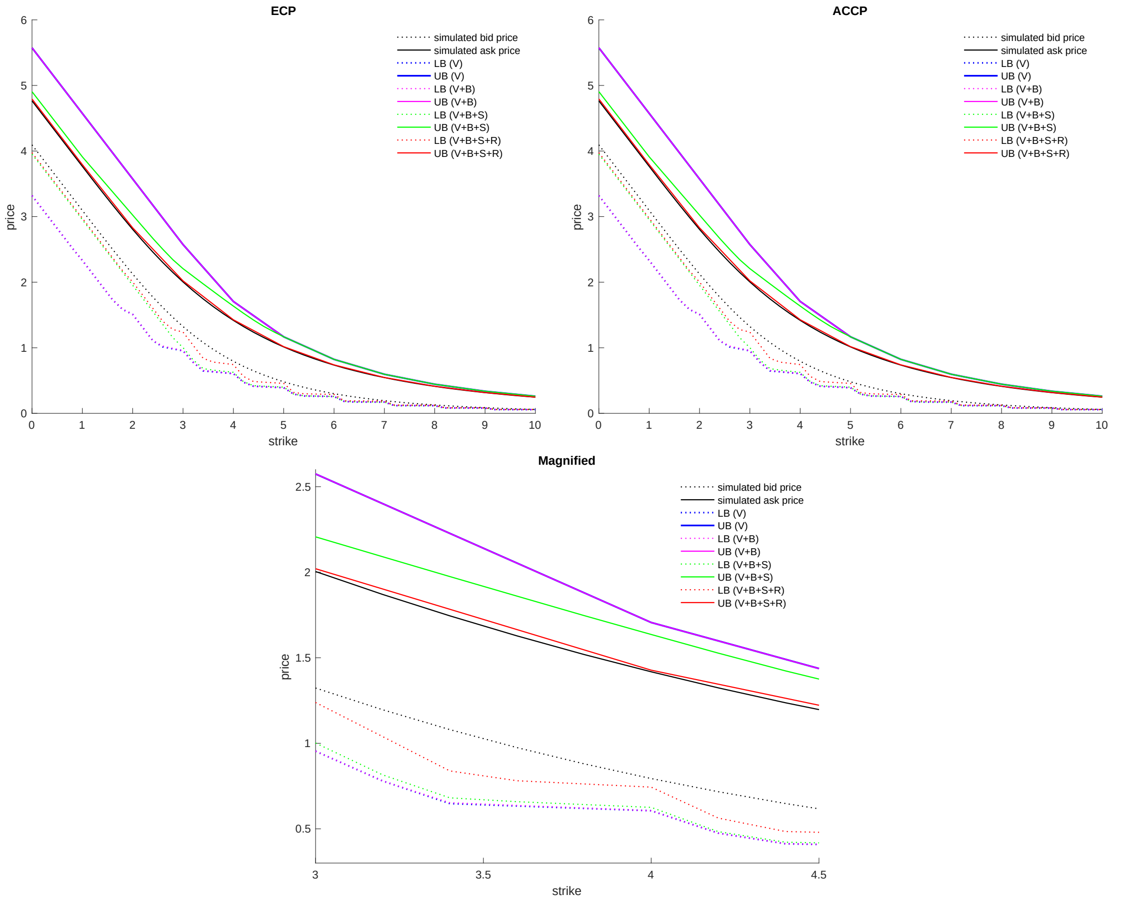

FIGURE 4.2. Experiment 2 – Model-free lower and upper price bounds of call-on-max options with strikes between 0 and 10 under 4 cases. The bottom panel shows a magnified version of a part of the top-right panel.

which is a t-copula with the same correlation matrix as the one introduced in Section 4.1, but with the degree of freedom parameter ν set to 10. Effectively, we have now created 4 market models, each made up of a combination of marginal distributions and dependence models. After specifying these models, we compute the bid and ask prices of the traded derivatives, as well as the bid and ask prices of the call-on-max option with different strikes, using the method described in the introduction of this section.

As in Experiment 1, we use the two algorithms with the same inputs. Figure 4.2 shows the computed lower and upper price bounds of the call-on-max option with different strikes, along with the bid and ask prices simulated via Monte Carlo from the same market model that generated the prices of traded derivatives. The simulated bid and ask prices represent the "tightest bounds" in this experiment.

One can observe that the price bounds in Cases V and V+B are identical. The price bounds in Cases V+B, V+B+S and V+B+S+R are still distinct. In Case V+B+S+R with all traded derivatives taken into consideration, the model-free upper bounds are very close to the simulated ask prices and the model-free lower bounds are very close to the simulated bid prices of the option. Again, the price bounds computed by the two algorithms are almost identical, and indeed we have checked that all of the absolute differences between the bounds computed by the two algorithms are below $\varepsilon = 0.001$.

4.3. **Experiment 3.** In this experiment, we consider a financial market with 60 assets ($d = 60$) which has bid-ask spread. We consider Setting 2 (Assumption 3.18), and let $\Omega = [0, 100]^{60}$. Our goal is to use Algorithm 1 and Algorithm 2 to compute the model-free lower and upper price bounds of a call-on-min option on the first 50 out of the 60 assets, with the strike price ranging from 0 to 1 with an

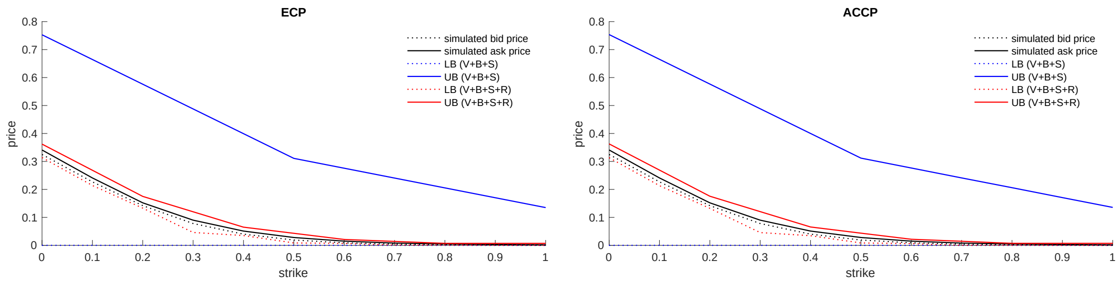

FIGURE 4.3. Experiment 3 – Model-free lower and upper price bounds of a call-on-min option with strikes between 0 and 1.

increment of 0.1. The purpose of this experiment is to demonstrate that Algorithm 1 and Algorithm 2 still work even when the number of assets is large.

A total of 400 financial derivatives are traded in the market ($m = 400$). These financial derivatives include:

- The 60 assets.
- 180 vanilla call options.
- 3 basket call options.
- 147 spread call options.
- 10 call-on-min options.

The market models we are considering are similar to those introduced in Section 4.1, where the marginal distributions of asset prices are log-normal, and the dependence model is specified as a t-copula with a correlation matrix which has a 3-factor model structure. For simplicity, we only consider Cases *V+B+S* and *V+B+S+R* in this experiment.

In Algorithm 1, we set $\varepsilon = 0.001, \tau = 1, \delta = 0.3$. We set $\underline{\phi}$ according to Remark 3.14. When computing $\phi(f)$, we set $\underline{\phi} = 0$, which is given by $0 \geq -\underline{f}$. When computing $\phi(-f)$, we set $\underline{\phi} = -\min\{\overline{\pi}_1, \dots, \overline{\pi}_{50}\}$, which follows from $x_1 \geq f, \dots, x_{50} \geq f$. In Algorithm 2, we set $\varepsilon = 0.001, \tau = 1, \gamma = 0.1, \zeta = 0.8, \delta = 0.3, \overline{c} = 100, \overline{\boldsymbol{y}} = 100 \cdot \mathbf{1}$. $\underline{\phi}$ is set to be the same as in Algorithm 1 and $\overline{\phi}$ is set according to Remark 3.22. When computing $\phi(f)$, we set $\overline{\phi} = \min\{\overline{\pi}_1, \dots, \overline{\pi}_{50}\}$ and $c^{\star(0)} + \langle \boldsymbol{y}^{\star(0)}, \boldsymbol{g} \rangle = x_{i_{\min}} \geq f$, where $i_{\min} := \arg\min_{1 \leq i \leq 50} \overline{\pi}_i$. When computing $\phi(-f)$, we set $\underline{\phi} = 0$ and $c^{\star(0)} + \langle \boldsymbol{y}^{\star(0)}, \boldsymbol{g} \rangle = 0 \geq -f$.

Remark 4.3. In order to prevent Algorithm 1 and Algorithm 2 from spending too much time on solving the MILP problem (3.14), we set an upper limit on the number of nodes explored by the BnB algorithm. This is fine for Algorithm 1 and for Algorithm 2 when $\overline{s}^{(r)} < 0$ in Line 20. However, in Line 20 of Algorithm 2, when the node limit of the BnB algorithm is reached and $\overline{s}^{(r)} \geq 0$, the algorithm might get stuck when Line 23 is not subsequently reached. To resolve this issue, we heuristically modify $\underline{s}^{(r)}$ to guarantee that Line 23 is subsequently reached. We observe that the BnB algorithm is almost always able to find the optimal integer feasible solution early on while spending most of the execution time refining the lower bound. Moreover, we empirically observe that the described scenario only occurred when $\phi(f)^{\mathrm{UB}} - \phi(f)^{\mathrm{LB}}$ is very small. Therefore, this heuristic adjustment usually does not affect the validity of Algorithm 2.

Figure 4.3 shows the computed lower and upper price bounds of the call-on-min option with different strikes, along with the bid and ask prices simulated via Monte Carlo from the same market model that generated the prices of traded derivatives. Notice that the lower price bounds in Case *V+B+S* are identically zero, showing that the traded vanilla, basket and spread options do not provide enough information for a non-trivial lower price bound of the call-on-min options, and that it is not possible to sub-replicate the payoff of a call-on-min option with these traded options. One the other hand, the inclusion of traded call-on-min options leads to non-trivial lower price bounds and much tighter no-arbitrage gap. This is similar to the observations from Experiments 1 and 2. Once again, the price

Algorithm	Problem	V+B+S	V+B+S+R
ECP	LP	4789	3339
	MILP	4789	3339
ACCP	LP	1639	1714
	MILP	1461	1574

TABLE 4.1. Experiment 3 – Total number of LP and MILP problems solved by the two algorithms.

bounds computed by the two algorithms are almost identical, and we have checked that all of the absolute differences between the bounds computed by the two algorithms are below $\varepsilon = 0.001$.

Table 4.1 shows the total number of LP and MILP problems solved throughout this experiment by the two algorithms. Recall that, as stated in Remark 3.14, the feasibility constraints returned from one run of Algorithm 1 or Algorithm 2 can be reused for the next run. As a result, most of the iterations occurred when pricing the first option, and later runs had significantly fewer iterations. The ACCP algorithm achieved convergence faster than the ECP algorithm in this experiment. Moreover, in the ACCP algorithm, the MILP problems were only approximately solved with relative gap tolerance $\zeta = 0.8$, as explained in Remark 3.22. As a result, the ACCP algorithm was much faster than the ECP algorithm in this experiment.

4.4. **Experiment 4.** In this experiment, we consider a market with 5 assets ($d = 5$) which has bid-ask spread. We consider Setting 1 (Assumption 3.6), in which $\Omega = \mathbb{R}^d_+$. The purpose of this experiment is to show the workings of Algorithm 0 and Algorithm 1 under Setting 1. The goal of this experiment is to use Algorithm 1 to compute the model-free lower and upper price bounds of best-of-call options with different strikes for each of the 5 assets. Specifically, the payoff function we are pricing is given by

$$f(\boldsymbol{x}) = (x_1 - \kappa_1)^+ \vee (x_2 - \kappa_2)^+ \vee (x_3 - \kappa_3)^+ \vee (x_4 - \kappa_4)^+ \vee (x_5 - \kappa_5)^+,$$

where $\kappa_1, \ldots, \kappa_5$ are randomly generated integers between 1 and 10. We randomly generate 10 such best-of-call options, and let their payoff functions be denoted by $f_1, \ldots, f_{10}$, respectively. The set of traded financial derivatives and the market models considered in this experiment are the same as in Experiment 2. In particular, we take into consideration the 5 assets and all vanilla, basket, spread and rainbow options that were used in Experiment 2.

In Algorithm 1, we set $\varepsilon = 0.001, \tau = 0.1, \delta = 0.3, \overline{\boldsymbol{x}} = (100, 100, 100, 100, 100)$. We set $\underline{\phi}$ according to Remark 3.14. When computing $\phi(f)$, we set $\underline{\phi} = 0$, which is given by $0 \geq -f$. When computing $\phi(-f)$, we set $\underline{\phi} = -\overline{\pi}_1 - \overline{\pi}_2 - \overline{\pi}_3 - \overline{\pi}_4 - \overline{\pi}_5$, which follows from $x_1 + x_2 + x_3 + x_4 + x_5 \geq f$.

In the following, we explain in detail the workings of Algorithm 0 to generate the radial constraints on $\boldsymbol{y}$, which are subsequently used by Algorithm 1. In this experiment, the slack function $s_{\boldsymbol{y}}(\boldsymbol{x})$, defined in Definition 3.7, contains a total of 337 terms, that is, when expressing $s_{\boldsymbol{y}}(\boldsymbol{x})$ by the form in (3.8), we have $K = 337$. Its radial function, $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z})$, contains 16 terms, that is, when expressing $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{x})$ by the form in (3.9), $\widetilde{K} = 16$. We also have $|\widetilde{\mathcal{I}}| = \prod_{k=1}^{\widetilde{K}} \widetilde{I}_k = 327,680$, thus the for-loop in Line 2 of Algorithm 0 was run 327,680 times. However, we do not actually need to check the condition (3.11) 327,680 times, since, if we can find both $\boldsymbol{v} \in A_{(i_k)}$ and $-\boldsymbol{v} \in A_{(i_k)}$, then we already know that the condition (3.11) holds. During the execution of Algorithm 0, the condition (3.11) was checked 160 times by the LP solver. In 58 of 160 times the condition held, and in 102 of 160 times the condition did not hold. At the end of the execution, we generated 510 linear inequality constraints with 966 auxiliary variables. These constraints were later used in Algorithm 1 to guarantee that $\inf_{\boldsymbol{x} \in \mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x}) > -\infty$ and the MILP problem in Line 12 has a solution.

Figure 4.4 shows the computed lower and upper price bounds of the 10 best-of-call options along with the bid and ask prices simulated via Monte Carlo from the same market model that generated the prices of traded derivatives. Similar to Experiment 2, the differences between the model-free lower bound and the simulated bid price and between the model-free upper bound and the simulated

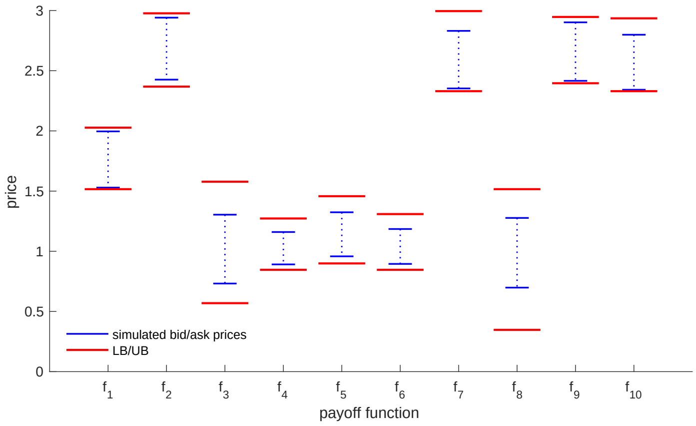

FIGURE 4.4. Experiment 4 – Model-free lower and upper price bounds of 10 different best-of-call options.

ask price are small, indicating that the market contains enough information to accurately price such options.

4.5. **Experiment 5.** Finally, in this experiment, we demonstrate the case where Assumption 2.1 does not hold. To do so, we consider the same set of traded financial derivatives as in Experiment 2 with their corresponding bid and ask prices, and include an additional mis-priced option in the market. The mis-priced option we choose is a call-on-max option, with the following payoff function denoted by $\widehat{g}$:

$$\widehat{g}(\boldsymbol{x}) := \big(x_2 \vee x_3 \vee x_4 - 1\big)^+.$$

In Experiment 2, we have already computed the model-free lower and upper price bounds for this call-on-max option. The model-free lower price bound is 2.9744 and the the model-free upper price bound is 3.7931.

In this experiment, we first set the ask price of the option as 2.9644, which is 0.01 less than the model-free lower bound, and set the bid price of the option as 0. Subsequently, we let $f = 0$ and run Algorithm 2 with the same settings as in Experiment 2. As expected, Line 37 is reached, where $c^\star$ and $\boldsymbol{y}^\star$ give the following non-negative portfolio (subject to rounding):

$$\begin{aligned}
c^\star + \langle \boldsymbol{y}^\star, \boldsymbol{g}\rangle &= 100\big(1 + (x_1 - x_3)^+ - x_1 \vee x_2 \vee x_3 \vee x_4 + \widehat{g}(\boldsymbol{x})\big)\\
&= \begin{cases} 100\big((1 - x_2 \vee x_3 \vee x_4)^+ + (x_1 - x_3)^+\big), & \text{if } x_1 \leq x_2 \vee x_3 \vee x_4,\\ 100\big((x_2 \vee x_3 \vee x_4 - 1)^+ - x_3 + 1\big), & \text{if } x_1 > x_2 \vee x_3 \vee x_4 \end{cases}\\
&\geq 0,
\end{aligned}$$

with $c^\star + \pi(\boldsymbol{y}) < 0$. Next, we set the bid price of the option as 3.8031, which is 0.01 more than the model-free upper bound, and set the ask price of the option as 100. When we run Algorithm 2 again, Line 37 is reached, and $c^\star$ and $\boldsymbol{y}^\star$ give the following non-negative portfolio (subject to rounding):

$$c^\star + \langle \boldsymbol{y}^\star, \boldsymbol{g}\rangle = 100\big((x_1 \vee x_2 \vee x_3 \vee x_4 - 1)^+ - \widehat{g}(\boldsymbol{x})\big) \geq 0,$$

with $c^\star + \pi(\boldsymbol{y}) < 0$. These results show that Assumption 2.1 is violated, and confirms Theorem 3.24(iv).

Appendix A. Proof of duality results

This appendix is dedicated to the proof of the fundamental theorem (Theorem 2.3) and the super-hedging duality (Theorem 2.4). Let us first prove the following two lemmata which are intermediate results.

Lemma A.1. *Denote the set of payoff functions that can be super-replicated by*

$$\mathcal{C} := \{\langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) : \boldsymbol{y} \in \mathbb{R}^m\} - \mathcal{L}^0_+, \tag{A.1}$$

where $\mathcal{L}^0_+ := \{h : \Omega \mapsto \mathbb{R} : h(\omega) \geq 0 \ \forall \omega \in \Omega\}$ *denotes the set of all measurable non-negative functions. Under Assumption 2.1, if* $(f_n)_{n\geq 1} \subset \mathcal{C}$ *and* $f_n \to f$ *point-wise, then* $f \in \mathcal{C}$.

Proof (adapted from Theorem 2.2 of [15]). Suppose $(f_n)_{n\geq 1} \subset \mathcal{C}$ and $f_n \to f$ point-wise. Then, there exist $\boldsymbol{y}_n \in \mathbb{R}^m, h_n \in \mathcal{L}^0_+$ such that $f_n = \langle \boldsymbol{y}_n, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_n) - h_n$, for $n \geq 1$. The goal is to find a subsequence $(\boldsymbol{y}_{n_k})_{k\geq 1}$ such that $\lim_{k\to\infty}\langle \boldsymbol{y}_{n_k}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_{n_k}) = \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y})$ point-wise for some $\boldsymbol{y} \in \mathbb{R}^m$. Given such a subsequence, for any $\omega \in \Omega$,

$$\begin{aligned}\langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle - \pi(\boldsymbol{y}) - f(\omega) &= \lim_{k\to\infty} \langle \boldsymbol{y}_{n_k}, \boldsymbol{g}(\omega)\rangle - \pi(\boldsymbol{y}_{n_k}) - f_{n_k}(\omega) \\ &= \lim_{k\to\infty} h_{n_k}(\omega) \geq 0,\end{aligned}$$

hence $f \in \mathcal{C}$.

Let us consider the following two separate cases.

Case 1: $\liminf_{n\to\infty} \|\boldsymbol{y}_n\| < \infty$. In this case, there exists a subsequence $(\boldsymbol{y}_{n_k})_{k\geq 1}$ that converges to some $\boldsymbol{y} \in \mathbb{R}^m$ by compactness, and $\lim_{k\to\infty}\langle \boldsymbol{y}_{n_k}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_{n_k}) = \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y})$ follows from the continuity of $\pi(\cdot)$.

Case 2: $\liminf_{n\to\infty} \|\boldsymbol{y}_n\| = \infty$. Define $\widetilde{\boldsymbol{y}}_n := \frac{\boldsymbol{y}_n}{1+\|\boldsymbol{y}_n\|}$. Then $\lim_{n\to\infty} \|\widetilde{\boldsymbol{y}}_n\| = 1$ and there exists a subsequence of $(\widetilde{\boldsymbol{y}}_n)_{n\geq 1}$, denoted again by $(\widetilde{\boldsymbol{y}}_n)_{n\geq 1}$ for simplicity, such that $\widetilde{\boldsymbol{y}}_n \to \widetilde{\boldsymbol{y}} \in \mathbb{R}^m$ as $n \to \infty$. Moreover, for all $\omega \in \Omega$,

$$\begin{aligned}\langle \widetilde{\boldsymbol{y}}, \boldsymbol{g}(\omega)\rangle - \pi(\widetilde{\boldsymbol{y}}) &= \lim_{n\to\infty} \langle \widetilde{\boldsymbol{y}}_n, \boldsymbol{g}(\omega)\rangle - \pi(\widetilde{\boldsymbol{y}}_n) \\ &= \lim_{n\to\infty} \frac{1}{1+\|\boldsymbol{y}_n\|}\left[\langle \boldsymbol{y}_n, \boldsymbol{g}(\omega)\rangle - \pi(\boldsymbol{y}_n)\right] \\ &\geq \lim_{n\to\infty} \frac{f_n(\omega)}{1+\|\boldsymbol{y}_n\|} \\ &= \lim_{n\to\infty} \frac{f_n(\omega) - f(\omega)}{1+\|\boldsymbol{y}_n\|} + \frac{f(\omega)}{1+\|\boldsymbol{y}_n\|} = 0,\end{aligned}$$

where the second equality follows from the positive homogeneity of $\pi(\cdot)$. By Assumption 2.1, we have

$$\langle \widetilde{\boldsymbol{y}}, \boldsymbol{g}\rangle - \pi(\widetilde{\boldsymbol{y}}) = 0. \tag{A.2}$$

The rest of the proof follows from induction, with the following induction hypothesis:

$$\left.\begin{matrix}\forall I \subset \{1,\dots,m\} \text{ and } |I| = l \\ [\boldsymbol{y}_n]_j = 0 \ \forall j \in I, \ \text{for } n \geq 1\end{matrix}\right\} \implies \begin{cases}(\boldsymbol{y}_n)_{n\geq 1} \text{ admits a subsequence } (\boldsymbol{y}_{n_k})_{k\geq 1} \text{ such that} \\ \exists \boldsymbol{y} \in \mathbb{R}^m, \lim_{k\to\infty}\langle \boldsymbol{y}_{n_k}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_{n_k}) = \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}).\end{cases} \tag{A.3}$$

The claim in (A.3) holds trivially when $l = m$. The goal is to show that (A.3) holds when $l = 0$.

Now, let $\widetilde{I} \subset \{1,\dots,m\}$, $|\widetilde{I}| = l - 1$, and suppose that $[\boldsymbol{y}_n]_j = 0$ for all $j \in \widetilde{I}$ and $n \geq 1$. If $\liminf_{n\to\infty} \|\boldsymbol{y}_n\| < \infty$, then the claim holds by Case 1 above. Therefore, we can assume that $\liminf_{n\to\infty} \|\boldsymbol{y}_n\| = \infty$. Let us first restrict $(\boldsymbol{y}_n)_{n\geq 1}$ to a subsequence such that $\operatorname{sign}(\boldsymbol{y}_{n_1}) = \operatorname{sign}(\boldsymbol{y}_{n_2}) = \dots$, where for $\boldsymbol{x} \in \mathbb{R}^m$, $\operatorname{sign}(\boldsymbol{x}) := (\operatorname{sign}([\boldsymbol{x}]_1), \dots, \operatorname{sign}([\boldsymbol{x}]_m))^\mathsf{T} \in \{-1, 1\}^m$ and for $x \in \mathbb{R}$,

$$\operatorname{sign}(x) := \begin{cases} 1, & \text{if } x \geq 0, \\ -1, & \text{otherwise.}\end{cases}$$

Such a subsequence exists since m is finite. Let this subsequence be denoted by $(\boldsymbol{y}_n)_{n\geq 1}$ again. By Case 2 and by setting $\widetilde{\boldsymbol{y}}_n := \frac{\boldsymbol{y}_n}{1+\|\boldsymbol{y}_n\|}$, we get a subsequence (still denoted by $(\widetilde{\boldsymbol{y}}_n)_{n\geq 1}$) such that $\widetilde{\boldsymbol{y}}_n \to \widetilde{\boldsymbol{y}}$ as $n \to \infty$. Note that we have $\operatorname{sign}(\boldsymbol{y}_n) = \operatorname{sign}(\widetilde{\boldsymbol{y}}_n)$ for all n. Moreover, (A.2) holds and for $j = 1, \ldots, m$,

$$\operatorname{sign}([\boldsymbol{y}_n]_j) = \operatorname{sign}([\widetilde{\boldsymbol{y}}_n]_j) = \operatorname{sign}([\widetilde{\boldsymbol{y}}]_j) \quad \text{for all } n, \quad \text{if } [\widetilde{\boldsymbol{y}}]_j \neq 0. \tag{A.4}$$

For every $n \geq 1$, let $\lambda_n = \min\{\gamma_{n,1}, \ldots, \gamma_{n,m}\}$, where

$$\gamma_{n,j} = \begin{cases} \frac{[\boldsymbol{y}_n]_j}{[\widetilde{\boldsymbol{y}}]_j}, & \text{if } [\widetilde{\boldsymbol{y}}]_j \neq 0, \\ +\infty, & \text{otherwise}, \end{cases}$$

for $j = 1, \ldots, m$. By (A.4), $\gamma_{n,j} \geq 0$ for all n and j. Since $\liminf_{n\to\infty} \|\boldsymbol{y}_n\| = \infty$, $\lambda_n > 0$ except for finitely many n, and we assume without loss of generality that $\lambda_n > 0$ for all $n \geq 1$. In addition, $\lambda_n < \infty$ for all $n \geq 1$ since $\|\widetilde{\boldsymbol{y}}\| = 1$. Let $\bar{\boldsymbol{y}}_n := \boldsymbol{y}_n - \lambda_n \widetilde{\boldsymbol{y}}$, and observe that for $j \in \{1, \ldots, m\} \setminus \widetilde{I}$,

$$[\bar{\boldsymbol{y}}_n]_j = [\boldsymbol{y}_n]_j - \lambda_n [\widetilde{\boldsymbol{y}}]_j \begin{cases} \geq 0, & \text{if } [\widetilde{\boldsymbol{y}}]_j \geq 0, \\ \leq 0, & \text{if } [\widetilde{\boldsymbol{y}}]_j < 0, \\ = 0, & \text{if } \lambda_n = \gamma_{n,j}. \end{cases}$$

It follows that for all n,

$$\operatorname{sign}([\boldsymbol{y}_n]_j) = \operatorname{sign}([\bar{\boldsymbol{y}}_n]_j) \quad \text{if } [\bar{\boldsymbol{y}}_n]_j \neq 0. \tag{A.5}$$

Combining (A.4) and (A.5), it follows that if $\pi(\boldsymbol{y}_n) = \langle \boldsymbol{y}_n, \boldsymbol{q}\rangle$ for $\boldsymbol{q} \in \Delta$, then $\pi(\bar{\boldsymbol{y}}_n) = \langle \bar{\boldsymbol{y}}_n, \boldsymbol{q}\rangle$, $\pi(\widetilde{\boldsymbol{y}}) = \langle \widetilde{\boldsymbol{y}}, \boldsymbol{q}\rangle$. Hence, $\pi(\bar{\boldsymbol{y}}_n) = \pi(\boldsymbol{y}_n) - \lambda_n \pi(\widetilde{\boldsymbol{y}})$. Since for all $j \in \widetilde{I}$, $[\bar{\boldsymbol{y}}_n]_j = 0$, it holds that $\bar{\boldsymbol{y}}_n$ has at least l zero-entries. Again, by extracting a subsequence, we can assume, without loss of generality, that the additional zero-entries in $\bar{\boldsymbol{y}}_n$ all occur at the same position, that is, there exists $j' \in \{1, \ldots, m\} \setminus \widetilde{I}$ such that $[\bar{\boldsymbol{y}}_n]_{j'} = 0$ for all $n \geq 1$. Now since $|\widetilde{I} \cup \{j'\}| = l$, by the induction hypothesis (A.3), $(\bar{\boldsymbol{y}}_n)_{n\geq 1}$ admits a subsequence $(\bar{\boldsymbol{y}}_{n_k})_{k\geq 1}$ such that $\exists \boldsymbol{y} \in \mathbb{R}^m$, $\lim_{k\to\infty} \langle \bar{\boldsymbol{y}}_{n_k}, \boldsymbol{g}\rangle - \pi(\bar{\boldsymbol{y}}_{n_k}) = \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y})$. By (A.2), it holds that for $n \geq 1$,

$$\langle \boldsymbol{y}_n, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_n) = \langle \boldsymbol{y}_n, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_n) - \lambda_n(\langle \widetilde{\boldsymbol{y}}, \boldsymbol{g}\rangle - \pi(\widetilde{\boldsymbol{y}})) = \langle \bar{\boldsymbol{y}}_n, \boldsymbol{g}\rangle - \pi(\bar{\boldsymbol{y}}_n),$$

hence $\lim_{k\to\infty} \langle \boldsymbol{y}_{n_k}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_{n_k}) = \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y})$. We have thus proved the induction hypothesis for the case where $|I| = l - 1$. By induction, the existence of a subsequence $(\boldsymbol{y}_{n_k})_{k\geq 1}$ such that $\lim_{k\to\infty} \langle \boldsymbol{y}_{n_k}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_{n_k}) = \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y})$ for some $\boldsymbol{y} \in \mathbb{R}^m$ is established. The proof is now complete. □

Lemma A.2. *For any $\nu \in \mathcal{P}(\Omega)$, let $\widetilde{\mathcal{P}}_\nu := \left\{\mu \in \mathcal{P}(\Omega) : \nu \ll \mu, \int_\Omega |f| + \sum_{j=1}^m |g_j| \mathrm{d}\mu < \infty\right\}$, and let $\Gamma_\nu := \left\{\int_\Omega \boldsymbol{g} \mathrm{d}\mu : \mu \in \widetilde{\mathcal{P}}_\nu\right\} \subset \mathbb{R}^m$. Under Assumption 2.1, $\Delta \cap \operatorname{relint}(\Gamma_\nu) \neq \emptyset$. In particular, $\Delta \cap \operatorname{relint}(\Gamma) \neq \emptyset$.*

Proof (partially adapted from Lemma 3.3 of [15]). Fix a $\nu \in \mathcal{P}(\Omega)$. Suppose, for the sake of contradiction, that $\Delta \cap \operatorname{relint}(\Gamma_\nu) = \emptyset$. Since both Δ and Γ_ν are convex, and Δ is polyhedral, by Theorem 20.2 of Rockafellar [60], there exists a hyperplane (namely, there exist $\boldsymbol{y} \in \mathbb{R}^n, \boldsymbol{y} \neq \boldsymbol{0}$ and $\alpha \in \mathbb{R}$)

$$H := \{\boldsymbol{w} \in \mathbb{R}^n : \langle \boldsymbol{y}, \boldsymbol{w}\rangle = \alpha\}$$

that separates Δ and Γ_ν properly and that $\Gamma_\nu \not\subseteq H$. Suppose, without loss of generality, that Δ is contained in the half-space $H_- := \{\boldsymbol{w} : \langle \boldsymbol{y}, \boldsymbol{w}\rangle \leq \alpha\}$, then

$$\pi(\boldsymbol{y}) = \sup_{\boldsymbol{q}\in\Delta} \langle \boldsymbol{y}, \boldsymbol{q}\rangle \leq \alpha.$$

On the other hand,

$$\int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle \mathrm{d}\mu \geq \alpha \geq \pi(\boldsymbol{y}), \quad \forall \mu \in \widetilde{\mathcal{P}}_\nu. \tag{A.6}$$

This implies that $\int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \mathrm{d}\mu \geq 0 \ \forall \mu \in \widetilde{\mathcal{P}}_\nu$. Suppose, for the sake of contradiction, that there exists $\omega \in \Omega$ such that $\langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle - \pi(\boldsymbol{y}) = \beta < 0$. Let $\widetilde{\nu} = (\nu + \delta_\omega)/2$, where δ_ω is the Dirac measure at ω. By Theorem VII.57 of Dellacherie and Meyer [27], there exists $\widehat{\nu} \in \mathcal{P}(\Omega)$ such that $\widehat{\nu} \sim \widetilde{\nu}$ and $\int_\Omega |f| + \sum_{j=1}^m |g_j| \mathrm{d}\widehat{\nu} < \infty$. Therefore, $\widehat{\nu} \in \widetilde{\mathcal{P}}_\nu$ and $\widehat{\nu}(\{\omega\}) > 0$. Now for $\varepsilon > 0$, define η_ε by

$$\frac{\mathrm{d}\eta_\varepsilon}{\mathrm{d}\widehat{\nu}} = \frac{\mathbb{1}_{\{\omega\}} + \varepsilon}{\widehat{\nu}(\{\omega\}) + \varepsilon}.$$

It is clear that $\eta_\varepsilon \sim \widehat{\nu}$ and $\eta_\varepsilon \in \widetilde{\mathcal{P}}_\nu$ for all $\varepsilon > 0$. Moreover,

$$\int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \mathrm{d}\eta_\varepsilon = \frac{\beta \widehat{\nu}(\{\omega\}) + \varepsilon \int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \mathrm{d}\widehat{\nu}}{\widehat{\nu}(\{\omega\}) + \varepsilon} \to \beta < 0$$

when $\varepsilon \to 0$. This implies that there exist $\varepsilon > 0$ and $\eta_\varepsilon \in \widetilde{\mathcal{P}}_\nu$ such that $\int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \mathrm{d}\eta_\varepsilon < 0$, which is a contradiction to (A.6). Therefore, $\langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle - \pi(\boldsymbol{y}) \geq 0$ for all $\omega \in \Omega$ and it follows from Assumption 2.1 that $\langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) = 0$. Hence, by (A.6),

$$\alpha \leq \inf_{\mu \in \widetilde{\mathcal{P}}_\nu} \int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle \mathrm{d}\mu \leq \sup_{\mu \in \widetilde{\mathcal{P}}_\nu} \int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle \mathrm{d}\mu \leq \sup_{\boldsymbol{q} \in \Delta} \langle \boldsymbol{y}, \boldsymbol{q}\rangle = \pi(\boldsymbol{y}) \leq \alpha,$$

and we conclude that $\int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle \mathrm{d}\mu = \alpha$ for all $\mu \in \widetilde{\mathcal{P}}_\nu$ and thus $\Gamma_\nu \subseteq H$, which is a contradiction to $\Gamma_\nu \nsubseteq H$. The last statement of Lemma A.2 follows from the fact that $\widetilde{\mathcal{P}} = \bigcup_{\nu \in \mathcal{P}(\Omega)} \widetilde{\mathcal{P}}_\nu$, and that $\Gamma = \bigcup_{\nu \in \mathcal{P}(\Omega)} \Gamma_\nu$. The proof is now complete. □

Proof of Theorem 2.3 *(adapted from Theorem 3.1 of [15]).* Suppose that Assumption 2.1 holds. We have by Lemma A.2 that for any $\nu \in \mathcal{P}(\Omega)$, there exists $\mu \in \mathcal{P}(\Omega)$ such that $\nu \ll \mu$ and $\int_\Omega \boldsymbol{g} \mathrm{d}\mu \in \Delta$, hence (ii) holds.

Conversely, suppose that Assumption 2.1 does not hold. Then, there exists $\boldsymbol{y} \in \mathbb{R}^m$ such that $\langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \geq 0$ and there exists $\omega \in \Omega$ such that $\langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle - \pi(\boldsymbol{y}) > 0$. With this $\boldsymbol{y}$ and any $\mu \in \mathcal{Q}$, it holds that

$$\int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle \mathrm{d}\mu \leq \sup_{\boldsymbol{q} \in \Delta} \langle \boldsymbol{y}, \boldsymbol{q}\rangle = \pi(\boldsymbol{y}),$$

implying $\int_\Omega \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \mathrm{d}\mu = 0$. Hence, $\mu(\{\omega\}) = 0$ and $\delta_\omega \not\ll \mu$ for all $\mu \in \mathcal{Q}$, implying that (ii) does not hold. The proof is now complete. □

Proof of Theorem 2.4 *(partially adapted from Theorem 3.4 of [15]).* Let us first prove statement (i). Suppose that $\phi(f) = -\infty$, then for all $n \geq 1$, there exists $\boldsymbol{y}_n \in \mathbb{R}^m$ such that $-n + \langle \boldsymbol{y}_n, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_n) \geq f$, which implies that

$$\langle \boldsymbol{y}_n, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_n) \geq f + n \geq \min\{f + n, 1\},$$

and thus $\min\{f + n, 1\} \in \mathcal{C}$ where $\mathcal{C}$ is the set of super-replicable payoff functions defined as in the statement of Lemma A.1. By Lemma A.1, $\lim_{n\to\infty} \min\{f + n, 1\} = 1 \in \mathcal{C}$, which is a contradiction to Assumption 2.1.

Next, let us prove statement (ii) by showing the existence of $\boldsymbol{y} \in \mathbb{R}^m$ such that $\phi(f) + \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \geq f$. If $\phi(f) = +\infty$, the claim holds trivially. If $\phi(f)$ is finite, then for all $n \geq 1$, there exists $\boldsymbol{y}_n \in \mathbb{R}^m$ such that $\phi(f) + \frac{1}{n} + \langle \boldsymbol{y}_n, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_n) \geq f$, which implies that $f - \phi(f) - \frac{1}{n} \in \mathcal{C}$. By Lemma A.1, $f - \phi(f) = \lim_{n\to\infty} f - \phi(f) - \frac{1}{n} \in \mathcal{C}$, hence there exists $\boldsymbol{y} \in \mathbb{R}^m$ such that $\phi(f) + \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \geq f$.

Finally, let us prove statement (iii), that is, the duality (2.8). Assume first that f is bounded from above. This part of the proof is based on the semi-infinite generalization of Fenchel's Duality Theorem by Borwein and Lewis [14]. Let $\mathcal{M}(\Omega)$ denote the set of finite signed Borel measures on Ω.

Let

$$\begin{aligned}\widetilde{\mathcal{M}} :=& \left\{\mu \in \mathcal{M}(\Omega) : \int_\Omega \left(\sum_{j=1}^m |g_j|\right) \mathrm{d}|\mu| < \infty\right\},\\ \widetilde{\mathcal{M}}^* :=& \left\{h : \Omega \mapsto \mathbb{R} \text{ is Borel measurable} : |h| \le \alpha\left(\sum_{j=1}^m |g_j| + 1\right) \text{ for some } \alpha > 0\right\},\end{aligned}$$

where $|\mu|$ denotes the total variation of the signed measure μ. For $\mu \in \widetilde{\mathcal{M}}$ and $h \in \widetilde{\mathcal{M}}^*$, define

$$\langle h, \mu\rangle := \int_\Omega h \mathrm{d}\mu < \infty.$$

$\widetilde{\mathcal{M}}$ is a locally convex topological vector space equipped with the $\sigma(\widetilde{\mathcal{M}}, \widetilde{\mathcal{M}}^*)$ topology. Let $\widetilde{\mathcal{M}}_+ := \{\mu \in \widetilde{\mathcal{M}} : \mu \text{ is a positive measure}\}$. For $\mu \in \widetilde{\mathcal{M}}$, let

$$F(\mu) := \begin{cases} \int_\Omega (-f) \mathrm{d}\mu, & \text{if } \mu \in \widetilde{\mathcal{M}}_+, \\ \infty, & \text{if } \mu \notin \widetilde{\mathcal{M}}_+, \end{cases} \tag{A.7}$$

which is convex and proper since $-f$ is bounded from below. For $h \in \widetilde{\mathcal{M}}^*$, let

$$\begin{aligned} F^*(h) :=& \sup_{\mu \in \widetilde{\mathcal{M}}} \{\langle h, \mu\rangle - F(\mu)\} \\ =& \sup_{\mu \in \widetilde{\mathcal{M}}_+} \left\{\int_\Omega h + f \mathrm{d}\mu\right\} \\ =& \begin{cases} 0 & \text{if } -h(\omega) \ge f(\omega)\ \forall \omega \in \Omega, \\ \infty & \text{if } \exists \omega \in \Omega, -h(\omega) < f(\omega), \end{cases} \end{aligned} \tag{A.8}$$

where the last equality holds because $\gamma\delta_\omega \in \widetilde{\mathcal{M}}_+$ for all $\omega \in \Omega$ and $\gamma > 0$. Let $G : \widetilde{\mathcal{M}} \mapsto \mathbb{R}^{m+1}$ be defined as follows,

$$G(\mu) := \left(\int_\Omega 1 \mathrm{d}\mu, \int_\Omega g_1 \mathrm{d}\mu, \ldots, \int_\Omega g_m \mathrm{d}\mu\right)^\mathsf{T}. \tag{A.9}$$

G is linear and $\sigma(\widetilde{\mathcal{M}}, \widetilde{\mathcal{M}}^*)$-continuous by definition. Let G^T be the adjoint of G, defined as

$$\begin{aligned} G^\mathsf{T} : \mathbb{R}^{m+1} &\mapsto \widetilde{\mathcal{M}}^* \\ \boldsymbol{\lambda} &\mapsto G^\mathsf{T}\boldsymbol{\lambda} = [\boldsymbol{\lambda}]_1 + \langle[\boldsymbol{\lambda}]_{2:m+1}, \boldsymbol{g}\rangle. \end{aligned}$$

Let us denote $[\boldsymbol{\lambda}]_1 \in \mathbb{R}$ by c and denote $[\boldsymbol{\lambda}]_{2:m+1} \in \mathbb{R}^m$ by $\boldsymbol{y}$. Thus,

$$G^\mathsf{T}\boldsymbol{\lambda} = c + \langle \boldsymbol{y}, \boldsymbol{g}\rangle. \tag{A.10}$$

Let $\widehat{\Delta} := \{1\} \times \Delta \subset \mathbb{R}^{m+1}$. Let $\chi_{\widehat{\Delta}}$ be the characteristic function of $\widehat{\Delta}$, that is,

$$\chi_{\widehat{\Delta}}(\boldsymbol{q}) := \begin{cases} 0, & \text{if } \boldsymbol{q} \in \widehat{\Delta}, \\ \infty, & \text{if } \boldsymbol{q} \notin \widehat{\Delta}. \end{cases} \tag{A.11}$$

Let $\chi^*_{\widehat{\Delta}}$ be the convex conjugate of $\chi_{\widehat{\Delta}}$, which is given by

$$\begin{aligned} \chi^*_{\widehat{\Delta}}(\boldsymbol{\lambda}) =& \sup_{\boldsymbol{q} \in \mathbb{R}^{m+1}} \left\{\langle \boldsymbol{\lambda}, \boldsymbol{q}\rangle - \chi_{\widehat{\Delta}}(\boldsymbol{q})\right\} \\ =& [\boldsymbol{\lambda}]_1 + \pi([\boldsymbol{\lambda}]_{2:m+1}) \\ =& c + \pi(\boldsymbol{y}). \end{aligned} \tag{A.12}$$

Let $\phi^D(f)$ denote the right-hand side of (2.8). By definitions (A.7), (A.9) and (A.11), $\phi^D(f)$ can be equivalently expressed as follows,

$$\begin{aligned}\phi^D(f) &:= \sup_{\mu\in\mathcal{Q}}\int_\Omega f\mathrm{d}\mu \\ &= -\inf_{\mu\in\widetilde{\mathcal{M}}_+}\left\{\int_\Omega(-f)\mathrm{d}\mu+\chi_{\widehat{\Delta}}(G(\mu))\right\} \\ &= -\inf_{\mu\in\widetilde{\mathcal{M}}}\left\{F(\mu)+\chi_{\widehat{\Delta}}(G(\mu))\right\}.\end{aligned} \tag{A.13}$$

It holds that $\chi_{\widehat{\Delta}}$ is convex and proper in $\mathbb{R}^{m+1}$ and F is convex and proper in $\widetilde{\mathcal{M}}$. By Theorem 4.2 of Borwein and Lewis [14], the following duality holds under the additional constraint qualification that $\operatorname{relint}(G(\operatorname{dom}F))\cap\operatorname{dom}\chi_{\widehat{\Delta}}\neq\emptyset$ (notice that $\chi_{\widehat{\Delta}}$ is a polyhedral function since its epigraph is a polyhedral set),

$$\begin{aligned}-\phi^D(f) &= \inf_{\mu\in\widetilde{\mathcal{M}}}\left\{F(\mu)+\chi_{\widehat{\Delta}}(G(\mu))\right\} \\ &= \sup_{\boldsymbol{\lambda}\in\mathbb{R}^{m+1}}\left\{-F^*(-G^\mathsf{T}\boldsymbol{\lambda})-\chi^*_{\widehat{\Delta}}(\boldsymbol{\lambda})\right\} \\ &= \sup\left\{-c-\pi(\boldsymbol{y}) : c\in\mathbb{R},\boldsymbol{y}\in\mathbb{R}^m, c+\langle\boldsymbol{y},\boldsymbol{g}\rangle\geq f\right\} \\ &= -\inf\left\{c+\pi(\boldsymbol{y}) : c\in\mathbb{R},\boldsymbol{y}\in\mathbb{R}^m, c+\langle\boldsymbol{y},\boldsymbol{g}\rangle\geq f\right\} = -\phi(f),\end{aligned} \tag{A.14}$$

where the third equality follows from definitions (A.8), (A.10) and (A.12). Hence, we only need to show that the constraint qualification $\operatorname{relint}(G(\operatorname{dom}F))\cap\operatorname{dom}\chi_{\widehat{\Delta}}\neq\emptyset$ holds. It holds that

$$\begin{aligned}\operatorname{relint}(G(\operatorname{dom}F))\cap\operatorname{dom}\chi_{\widehat{\Delta}} &= \operatorname{relint}(G(\operatorname{dom}F))\cap(\{1\}\times\mathbb{R}^m)\cap(\mathbb{R}\times\Delta) \\ &= \operatorname{relint}(G(\operatorname{dom}F)\cap(\{1\}\times\mathbb{R}^m))\cap(\mathbb{R}\times\Delta).\end{aligned}$$

Since

$$\begin{aligned}\operatorname{dom}F &= \left\{\mu\in\widetilde{\mathcal{M}}_+ : \int_\Omega(-f)\mathrm{d}\mu<\infty\right\} \\ &\supseteq \left\{\mu\in\mathcal{M}(\Omega):\mu\geq 0, \int_\Omega |f|+\textstyle\sum_{j=1}^m|g_j|\mathrm{d}\mu<\infty\right\},\end{aligned}$$

we have

$$\begin{aligned}G(\operatorname{dom}F)\cap(\{1\}\times\mathbb{R}^m) &\supseteq G\left(\left\{\mu\in\mathcal{M}(\Omega):\mu\geq 0, \mu(\Omega)=1, \int_\Omega |f|+\textstyle\sum_{j=1}^m|g_j|\mathrm{d}\mu<\infty\right\}\right) \\ &= G(\widetilde{\mathcal{P}}) = \{1\}\times\Gamma.\end{aligned}$$

Since $\operatorname{relint}(\{1\}\times\Gamma)\cap(\mathbb{R}\times\Delta) = \{1\}\times(\operatorname{relint}(\Gamma)\cap\Delta)$, which is non-empty by Lemma A.2, we have that $\operatorname{relint}(G(\operatorname{dom}F))\cap\operatorname{dom}\chi_{\widehat{\Delta}}\supseteq\operatorname{relint}(\{1\}\times\Gamma)\cap(\mathbb{R}\times\Delta)\neq\emptyset$ and the constraint qualification holds.

Now suppose that f is unbounded from above. Then it holds that for all $n\geq 1$,

$$\phi(\min\{f,n\}) = \sup_{\mu\in\mathcal{Q}}\int_\Omega\min\{f,n\}\mathrm{d}\mu. \tag{A.15}$$

On the one hand,

$$\limsup_{n\to\infty}\sup_{\mu\in\mathcal{Q}}\int_\Omega\min\{f,n\}\mathrm{d}\mu\leq\sup_{\mu\in\mathcal{Q}}\int_\Omega f\mathrm{d}\mu. \tag{A.16}$$

On the other hand, $\limsup_{n\to\infty}\phi(\min\{f,n\})\leq\phi(f)$. Let $\limsup_{n\to\infty}\phi(\min\{f,n\})=\alpha\in\mathbb{R}\cup\{\infty\}$. If $\alpha=\infty$, then by (A.16) and (A.15),

$$\sup_{\mu\in\mathcal{Q}}\int_\Omega f\mathrm{d}\mu\geq\limsup_{n\to\infty}\sup_{\mu\in\mathcal{Q}}\int_\Omega\min\{f,n\}\mathrm{d}\mu=\limsup_{n\to\infty}\phi(\min\{f,n\})=\infty,$$

and both sides of (2.8) are equal to ∞. If $\alpha < \infty$, then there exists $\boldsymbol{y}_n \in \mathbb{R}^m$ such that $\alpha + \langle \boldsymbol{y}_n, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_n) \geq \min\{f, n\}$ by statement (ii) of Theorem 2.4. It follows that $\min\{f, n\} - \alpha \in \mathcal{C}$ hence $f - \alpha = \lim_{n\to\infty} \min\{f, n\} - \alpha \in \mathcal{C}$ by Lemma A.1. Therefore $\phi(f) \leq \alpha$ and

$$\phi(f) = \alpha = \limsup_{n\to\infty} \phi(\min\{f, n\}) = \limsup_{n\to\infty} \sup_{\mu\in\mathcal{Q}} \int_\Omega \min\{f, n\}\mathrm{d}\mu \leq \sup_{\mu\in\mathcal{Q}} \int_\Omega f\mathrm{d}\mu$$

by (A.16) and (A.15). The reverse direction $\phi(f) \geq \sup_{\mu\in\mathcal{Q}} \int_\Omega f\mathrm{d}\mu$ is easier to verify. Suppose that $\phi(f) < \infty$, since otherwise $\phi(f) \geq \sup_{\mu\in\mathcal{Q}} \int_\Omega f\mathrm{d}\mu$ holds trivially. For any $c \in \mathbb{R}, \boldsymbol{y} \in \mathbb{R}^m$ such that $c + \langle \boldsymbol{y}, \boldsymbol{g}\rangle \geq f$, and any $\mu \in \mathcal{Q}$, the following holds:

$$c + \pi(\boldsymbol{y}) = c + \sup_{\boldsymbol{q}\in\Delta}\langle \boldsymbol{y}, \boldsymbol{q}\rangle \geq \int_\Omega c + \langle \boldsymbol{y}, \boldsymbol{g}\rangle\mathrm{d}\mu \geq \int_\Omega f\mathrm{d}\mu. \tag{A.17}$$

Taking the infimum over all $c \in \mathbb{R}, \boldsymbol{y} \in \mathbb{R}^m$ such that $c + \langle \boldsymbol{y}, \boldsymbol{g}\rangle \geq f$ and taking the supremum over all $\mu \in \mathcal{Q}$ gives $\phi(f) \geq \sup_{\mu\in\mathcal{Q}} \int_\Omega f\mathrm{d}\mu$. It follows that (2.8) holds. The proof is now complete. □

Proof of Proposition 2.9. Suppose that Assumption 2.1 does not hold, and there exist $\boldsymbol{y} \in \mathbb{R}^m$ and $\widehat{\omega} \in \Omega$ such that $\langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}) \geq 0$ and $\langle \boldsymbol{y}, \boldsymbol{g}(\widehat{\omega})\rangle - \pi(\boldsymbol{y}) = \alpha > 0$. By the continuity of $\boldsymbol{g}$, there exists an open set $E \subset \Omega$ such that $\widehat{\omega} \in E$ and $\langle \boldsymbol{y}, \boldsymbol{g}(\omega)\rangle - \pi(\boldsymbol{y}) > \frac{\alpha}{2}$ for all $\omega \in E$. Since $\widehat{\mu}$ is equivalent to the Lebesgue measure on Ω, we have $\widehat{\mu}(E) > 0$, and thus

$$\int \langle \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y})\mathrm{d}\widehat{\mu} \geq \int \frac{\alpha}{2}\mathbb{1}_E\mathrm{d}\widehat{\mu} = \frac{\alpha}{2}\widehat{\mu}(E) > 0.$$

This implies that $\int \langle \boldsymbol{y}, \boldsymbol{g}\rangle\mathrm{d}\widehat{\mu} > \pi(\boldsymbol{y})$. However, since $\widehat{\mu} \in \mathcal{Q}$, we also have $\int \boldsymbol{g}\mathrm{d}\widehat{\mu} \in \Delta$, therefore $\int \langle \boldsymbol{y}, \boldsymbol{g}\rangle\mathrm{d}\widehat{\mu} \leq \sup_{\boldsymbol{q}\in\Delta}\langle \boldsymbol{y}, \boldsymbol{q}\rangle = \pi(\boldsymbol{y})$, which is a contradiction. The proof is now complete. □

Appendix B. Proof of results related to the numerical methods

Proof of Lemma 3.3. Property (i) is immediately clear from Definition 3.1. To prove the property (ii), suppose that h has the form in (3.4). Notice that each term in (3.4) can be written as follows,

$$\xi_k \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}\rangle + b_{k,i} : 1 \leq i \leq I_k\right\} = \begin{cases} \langle \xi_k \boldsymbol{a}_{k,1}, \boldsymbol{x}\rangle + \xi_k b_{k,1}, & \text{if } \boldsymbol{x} \in \Omega_{k,1}, \\ \quad\vdots & \quad\vdots \\ \langle \xi_k \boldsymbol{a}_{k,I_k}, \boldsymbol{x}\rangle + \xi_k b_{k,I_k}, & \text{if } \boldsymbol{x} \in \Omega_{k,I_k}, \end{cases}$$

where $\Omega_{k,i} := \{\boldsymbol{x} \in \mathbb{R}^d_+ : \langle \boldsymbol{a}_{k,i}, \boldsymbol{x}\rangle + b_{k,i} \geq \langle \boldsymbol{a}_{k,i'}, \boldsymbol{x}\rangle + b_{k,i'}\ \forall 1 \leq i' \leq I_k\}$ is a closed polyhedron (if non-empty) in $\mathbb{R}^d_+$, and $\bigcup_{i=1}^{I_k} \Omega_{k,i} = \mathbb{R}^d_+$. Moreover, $\mathrm{int}(\Omega_{k,i}) = \{\boldsymbol{x} \in \mathbb{R}^d_+ : \langle \boldsymbol{a}_{k,i}, \boldsymbol{x}\rangle + b_{k,i} > \langle \boldsymbol{a}_{k,i'}, \boldsymbol{x}\rangle + b_{k,i'}\ \forall 1 \leq i' \leq I_k\}$. Thus, for $i \neq i'$, $\mathrm{int}(\Omega_{k,i}) \cap \mathrm{int}(\Omega_{k,i'}) = \emptyset$. If $\boldsymbol{x} \in \Omega_{k,i} \cap \Omega_{k,i'}$, by the definition of $\Omega_{k,i}$ and $\Omega_{k,i'}$, $\langle \xi_k \boldsymbol{a}_{k,i}, \boldsymbol{x}\rangle + \xi_k b_{k,i} = \langle \xi_k \boldsymbol{a}_{k,i'}, \boldsymbol{x}\rangle + \xi_k b_{k,i'}$. Let $\mathcal{I} := \{(i_k)_{k=1:K} : 1 \leq i_k \leq I_k\}$. h can be decomposed into $|\mathcal{I}| = \prod_{k=1}^K I_k$ cases, and expressed by the following local representation (note that for some cases the corresponding set can be empty):

$$h(\boldsymbol{x}) = \begin{cases} \left\langle \sum_{k=1}^K \xi_k \boldsymbol{a}_{k,1}, \boldsymbol{x}\right\rangle + \sum_{k=1}^K \xi_k b_{k,1}, & \text{if } \boldsymbol{x} \in \bigcap_{k=1}^K \Omega_{k,1}, \\ \quad\vdots & \quad\vdots \\ \left\langle \sum_{k=1}^K \xi_k \boldsymbol{a}_{k,i_k}, \boldsymbol{x}\right\rangle + \sum_{k=1}^K \xi_k b_{k,i_k}, & \text{if } \boldsymbol{x} \in \bigcap_{k=1}^K \Omega_{k,i_k}, \\ \quad\vdots & \quad\vdots \\ \left\langle \sum_{k=1}^K \xi_k \boldsymbol{a}_{k,I_k}, \boldsymbol{x}\right\rangle + \sum_{k=1}^K \xi_k b_{k,I_k}, & \text{if } \boldsymbol{x} \in \bigcap_{k=1}^K \Omega_{k,I_k}. \end{cases} \tag{B.1}$$

Let $\Omega_{(i_k)} := \bigcap_{k=1}^K \Omega_{k,i_k}$. It is straightforward to verify that for each $(i_k) \in \mathcal{I}$, $\Omega_{(i_k)}$ is a closed polyhedron (if non-empty), and $\bigcup_{(i_k)\in\mathcal{I}} \Omega_{(i_k)} = \mathbb{R}^d_+$. Let $(i_k) \in \mathcal{I}, (i'_k) \in \mathcal{I}$. If $i_j \neq i'_j$ for some $j \in \{1, \ldots, K\}$, then $\mathrm{int}(\Omega_{(i_k)}) \cap \mathrm{int}(\Omega_{(i'_k)}) \subset \mathrm{int}(\Omega_{j,i_j}) \cap \mathrm{int}(\Omega_{j,i'_j}) = \emptyset$. If $\boldsymbol{x} \in \Omega_{(i_k)} \cap \Omega_{(i'_k)}$, then

$$\langle \xi_k \boldsymbol{a}_{k,i_k}, \boldsymbol{x}\rangle + \xi_k b_{k,i_k} = \left\langle \xi_k \boldsymbol{a}_{k,i'_k}, \boldsymbol{x}\right\rangle + \xi_k b_{k,i'_k}$$

for all $k = 1, \dots, K$, thus

$$\left\langle \sum_{k=1}^{K} \xi_k \boldsymbol{a}_{k,i_k}, \boldsymbol{x} \right\rangle + \sum_{k=1}^{K} \xi_k b_{k,i_k} = \left\langle \sum_{k=1}^{K} \xi_k \boldsymbol{a}_{k,i'_k}, \boldsymbol{x} \right\rangle + \sum_{k=1}^{K} \xi_k b_{k,i'_k}.$$

We have completed the proof of property (ii).

To prove property (iii), we divide the collection of all polyhedrons in the representation (3.4) $\mathcal{O} := \{\Omega_j : j = 1, \dots, J\}$ into $\mathcal{N} := \{\Omega_j \in \mathcal{O} : \mathrm{int}(\Omega_j) \neq \emptyset\}$ and $\mathcal{O} \setminus \mathcal{N}$. We claim that $\bigcup_{O \in \mathcal{N}} O = \mathbb{R}^d_+$. Suppose there exists $\boldsymbol{x} \in \mathbb{R}^d_+ \setminus \left(\bigcup_{O \in \mathcal{N}} O\right)$. $\bigcup_{O \in \mathcal{N}} O$ is closed since it is a finite union of closed sets. By property (ii), $\mathbb{R}^d_+ \setminus \left(\bigcup_{O \in \mathcal{N}} O\right) = \bigcup_{O \in \mathcal{O} \setminus \mathcal{N}} O$. Therefore, there is an open set E, such that $\boldsymbol{x} \in E \subset \bigcup_{O \in \mathcal{O} \setminus \mathcal{N}} O$. This is a contradiction to the fact that $\bigcup_{O \in \mathcal{O} \setminus \mathcal{N}} O$ has empty interior. Hence, in the local representation (3.5), it is sufficient to only enumerate over those polyhedrons with non-empty interior, and we have proved property (iii). The proof is now complete. □

Proof of Proposition 3.5. Let us first prove statement (i). Suppose that h has the representation (3.4). Notice that for fixed $\boldsymbol{x}_0 \in \mathbb{R}^d_+$ and $\boldsymbol{z} \in \mathbb{R}^d_+$, $\gamma_k(t) := \max\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}_0 + t\boldsymbol{z}\rangle + b_{k,i} : 1 \le i \le I_k\}$ is a CPWA function in t defined on $\mathbb{R}_+$. Hence, for $\bar{t}_k > 0$ large enough, γ_k is an affine function on $[\bar{t}_k, \infty)$. Hence, for all $t \ge \bar{t}_k$,

$$\gamma_k(t) = \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}_0 + \bar{t}_k \boldsymbol{z}\rangle + b_{k,i} : 1 \le i \le I_k\right\} + (t - \bar{t}_k) \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{z}\rangle : 1 \le i \le I_k\right\}.$$

Therefore, for $\bar{t} := \max\{\bar{t}_1, \dots, \bar{t}_K\}$ and all $t \ge \bar{t}$,

$$\begin{aligned} h(\boldsymbol{x}_0 + t\boldsymbol{z}) &= \sum_{k=1}^{K} \xi_k \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}_0 + \bar{t}\boldsymbol{z}\rangle + b_{k,i} : 1 \le i \le I_k\right\} \\ &\quad + (t - \bar{t}) \sum_{k=1}^{K} \xi_k \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{z}\rangle : 1 \le i \le I_k\right\} \\ &= \sum_{k=1}^{K} \xi_k \max\left\{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}_0 + \bar{t}\boldsymbol{z}\rangle + b_{k,i} : 1 \le i \le I_k\right\} + (t - \bar{t})\widetilde{h}(\boldsymbol{z}). \end{aligned} \tag{B.2}$$

Suppose first that $\inf_{\boldsymbol{x} \in \mathbb{R}^d_+} h(\boldsymbol{x}) > -\infty$, that is, there exists $\alpha \in \mathbb{R}$ such that $h(\boldsymbol{x}) \ge \alpha$ for all $\boldsymbol{x} \in \mathbb{R}^d_+$. Thus, $h(\boldsymbol{x}_0 + t\boldsymbol{z}) \ge \alpha$ for all $t \ge 0$, and, because of (B.2), $\widetilde{h}(\boldsymbol{z}) \ge 0$. The same reasoning applies for every $\boldsymbol{z} \in \mathbb{R}^d_+$, and thus $\widetilde{h}(\boldsymbol{z}) \ge 0$ for all $\boldsymbol{z} \in \mathbb{R}^d_+$.

Conversely, suppose that $\inf_{\boldsymbol{x} \in \mathbb{R}^d_+} h(\boldsymbol{x}) = -\infty$. By Lemma 3.3(ii), h admits the following local representation,

$$h(\boldsymbol{x}) = \begin{cases} \langle \boldsymbol{a}_1, \boldsymbol{x}\rangle + b_1, & \text{if } \boldsymbol{x} \in \Omega_1, \\ \quad\vdots & \quad\vdots \\ \langle \boldsymbol{a}_J, \boldsymbol{x}\rangle + b_J, & \text{if } \boldsymbol{x} \in \Omega_J, \end{cases} \tag{B.3}$$

hence there exists a $j \in \{1, \dots, J\}$ such that $\inf\{\langle \boldsymbol{a}_j, \boldsymbol{x}\rangle : \boldsymbol{x} \in \Omega_j\} = -\infty$. By well-known results about polyhedrons (see *e.g.* Theorem 19.1 of Rockafellar [60]), the polyhedron Ω_j can be represented by its finite number of extreme points and extreme directions, that is,

$$\Omega_j = \left\{ \sum_{\boldsymbol{v} \in V} \lambda_{\boldsymbol{v}} \boldsymbol{v} + \sum_{\boldsymbol{z} \in D} \zeta_{\boldsymbol{z}} \boldsymbol{z} : \lambda_{\boldsymbol{v}} \ge 0 \; \forall \boldsymbol{v} \in V, \sum_{\boldsymbol{v} \in V} \lambda_{\boldsymbol{v}} = 1, \zeta_{\boldsymbol{z}} \ge 0 \; \forall \boldsymbol{z} \in D \right\}, \tag{B.4}$$

where $V \subset \mathbb{R}^d_+$ is the set of extreme points, $|V| < \infty$, $D \subset \mathbb{R}^d_+$ is the set of extreme directions, $|D| < \infty$. Given this representation, we have

$$\begin{aligned} &\inf\{\langle \boldsymbol{a}_j, \boldsymbol{x}\rangle : \boldsymbol{x} \in \Omega_j\} \\ &= \inf\left\{ \sum_{\boldsymbol{v} \in V} \lambda_{\boldsymbol{v}} \langle \boldsymbol{a}_j, \boldsymbol{v}\rangle + \sum_{\boldsymbol{z} \in D} \zeta_{\boldsymbol{z}} \langle \boldsymbol{a}_j, \boldsymbol{z}\rangle : \lambda_{\boldsymbol{v}} \ge 0 \; \forall \boldsymbol{v} \in V, \sum_{\boldsymbol{v} \in V} \lambda_{\boldsymbol{v}} = 1, \zeta_{\boldsymbol{z}} \ge 0 \; \forall \boldsymbol{z} \in D \right\} \end{aligned}$$

thus,

$$\inf\{\langle \boldsymbol{a}_j, \boldsymbol{x}\rangle : \boldsymbol{x} \in \Omega_j\} = -\infty \quad \Longrightarrow \quad \exists \boldsymbol{z}_0 \in D \text{ such that } \langle \boldsymbol{a}_j, \boldsymbol{z}_0\rangle < 0. \tag{B.5}$$

By the representation (B.4), for fixed $\boldsymbol{x}_0 \in \Omega_j$ and $\boldsymbol{z}_0$ given by (B.5), $\boldsymbol{x}_0 + t\boldsymbol{z}_0 \in \Omega_j$ for all $t \geq 0$ as $\boldsymbol{z}_0$ is an extreme direction. Thus, by (B.3), for all $t \geq 0$,

$$h(\boldsymbol{x}_0 + t\boldsymbol{z}_0) = \langle \boldsymbol{a}_j, \boldsymbol{x}_0\rangle + b_j + t\langle \boldsymbol{a}_j, \boldsymbol{z}_0\rangle.$$

Since $\langle \boldsymbol{a}_j, \boldsymbol{z}_0\rangle < 0$, we have $\lim_{t\to\infty} h(\boldsymbol{x}_0+t\boldsymbol{z}_0) = -\infty$. By (B.2), we can conclude that $\widetilde{h}(\boldsymbol{z}_0) < 0$. This settles *(i)*.

To prove statement (ii), suppose that $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} h(\boldsymbol{x}) > -\infty$. Again by Lemma 3.3(ii), h admits the local representation in (B.3), hence $\inf_{\boldsymbol{x}\in\Omega_j} h(\boldsymbol{x}) > -\infty$ for $j = 1, \ldots, J$. For each j, Ω_j can be decomposed as in (B.4) into the set of extreme points V and the set of extreme directions D. By (B.4) and the fact that $\inf_{\boldsymbol{x}\in\Omega_j} h(\boldsymbol{x}) > -\infty$, $\langle \boldsymbol{a}_j, \boldsymbol{z}\rangle \geq 0$ for all $\boldsymbol{z} \in D$. Hence, there exists $\boldsymbol{x}^\star_j \in V$ such that $h(\boldsymbol{x}^\star_j) = \inf_{\boldsymbol{x}\in\Omega_j} h(\boldsymbol{x})$. Therefore, there exists $\boldsymbol{x}^\star \in \mathbb{R}^d_+$ such that $h(\boldsymbol{x}^\star) = \inf_{\boldsymbol{x}\in\mathbb{R}^d_+} h(\boldsymbol{x})$. This settles *(ii)*. The proof is now complete. □

B.1. **Proof of results in Setting 1.**

B.1.1. *Proof of auxiliary results.*

Proof of Proposition 3.9. Statement (i) follows directly from Assumption 3.6, Definition 3.7, and Lemma 3.3(i).

To prove statement (ii), let us explicitly express g_j for $j = 1, \ldots, m$ and f as CPWA functions as in (3.4): for $j = 1, \ldots, m$,

$$g_j(\boldsymbol{x}) = \sum_{k=1}^{K^{g_j}} \xi^{g_j}_k \max\left\{\langle \boldsymbol{a}^{g_j}_{k,i}, \boldsymbol{x}\rangle + b^{g_j}_{k,i} : 1 \leq i \leq I^{g_j}_k\right\}, \tag{B.6}$$

where $K^{g_j} \in \mathbb{N}$, $I^{g_j}_k \in \mathbb{N}$ for $k = 1, \ldots, K^{g_j}$, $\boldsymbol{a}^{g_j}_{k,i} \in \mathbb{R}^d$, $b^{g_j}_{k,i} \in \mathbb{R}$, $\xi^{g_j}_k \in \{-1, 1\}$, for $i = 1, \ldots, I^{g_j}_k, k = 1, \ldots, K^{g_j}$,

$$f(\boldsymbol{x}) = \sum_{k=1}^{K^f} \xi^f_k \max\left\{\langle \boldsymbol{a}^f_{k,i}, \boldsymbol{x}\rangle + b^f_{k,i} : 1 \leq i \leq I^f_k\right\}, \tag{B.7}$$

where $K^f \in \mathbb{N}$, $I^f_k \in \mathbb{N}$ for $k = 1, \ldots, K^f$, $\boldsymbol{a}^f_{k,i} \in \mathbb{R}^d$, $b^f_{k,i} \in \mathbb{R}$, $\xi^f_k \in \{-1, 1\}$, for $i = 1, \ldots, I^f_k, k = 1, \ldots, K^f$. By Definition 3.7 and (B.6), (B.7),

$$\begin{aligned}
s_{\boldsymbol{y}}(\boldsymbol{x}) &= \sum_{j=1}^m y_j g_j - f \\
&= \sum_{j=1}^m \sum_{k=1}^{K^{g_j}} \xi^{g_j}_k y_j \max\left\{\langle \boldsymbol{a}^{g_j}_{k,i}, \boldsymbol{x}\rangle + b^{g_j}_{k,i} : 1 \leq i \leq I^{g_j}_k\right\} \\
&\quad - \sum_{k=1}^{K^f} \xi^f_k \max\left\{\langle \boldsymbol{a}^f_{k,i}, \boldsymbol{x}\rangle + b^f_{k,i} : 1 \leq i \leq I^f_k\right\},
\end{aligned}$$

which has the desired form in (3.8). Notice that one may extract positive coefficients out of $\max\{\cdot\}$ and combine terms to simplify this representation. The representation (3.9) follows directly from (3.8) and Definition 3.4. The number of terms in (3.9) could be different from the number of terms in (3.8) due to possible simplification. We have now proved statement (ii).

Statement (iii) is a direct consequence of Proposition 3.5.

To prove statement (iv), let us fix a $\boldsymbol{y}$ such that $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x}) > -\infty$ and express $s_{\boldsymbol{y}}(\boldsymbol{x})$ by (3.8). Let $\mathcal{I} := \{(i_k)_{k=1:K} : 1 \leq i_k \leq I_k\}$. Same as in the proof of Lemma 3.3(ii), $s_{\boldsymbol{y}}(\boldsymbol{x})$ has a local representation as in (B.1), with ξ_k replaced by $\langle \boldsymbol{y}, \boldsymbol{w}_k\rangle + z_k$, over the finite partition $\mathfrak{C} :=$

$\left\{\bigcap_{k=1}^{K} \Omega_{k,i_k} : (i_k) \in \mathcal{I}\right\}$ of $\mathbb{R}^d_+$, where $\Omega_{k,i} := \{\boldsymbol{x} \in \mathbb{R}^d_+ : \langle \boldsymbol{a}_{k,i}, \boldsymbol{x}\rangle + b_{k,i} \geq \langle \boldsymbol{a}_{k,i'}, \boldsymbol{x}\rangle + b_{k,i'}\ \forall 1 \leq i' \leq I_k\}$. Notice that the partition $\mathfrak{C}$ does not depend on the value of $\boldsymbol{y}$, and each $C \in \mathfrak{C}$ is a polyhedron. Let $X^\star := \bigcup_{C\in\mathfrak{C}} V(C)$, where $V(C)$ denotes the set of extreme points of the polyhedron C given in (B.4). Thus, $|X^\star| < \infty$, and by (B.4), for each $C \in \mathfrak{C}$, there exists $\boldsymbol{x}^\star \in V(C)$ (depending on $\boldsymbol{y}$) such that $s_{\boldsymbol{y}}(\boldsymbol{x}^\star) = \inf_{\boldsymbol{x}\in C} s_{\boldsymbol{y}}(\boldsymbol{x})$. Since $X^\star$ does not depend on $\boldsymbol{y}$, it has the required properties. The proof is now complete. □

Proof of Proposition 3.10. Let us first prove statement (i). Since

$$\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) = \sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k) \max\left\{\langle \widetilde{\boldsymbol{a}}_{k,i}, \boldsymbol{z}\rangle : 1 \leq i \leq \widetilde{I}_k\right\},$$

we have for fixed $(i_k) \in \widetilde{\mathcal{I}}$ and all

$$\boldsymbol{z} \in \left\{\boldsymbol{z} \in \mathbb{R}^d_+ : \langle \widetilde{\boldsymbol{a}}_{k,i_k}, \boldsymbol{z}\rangle \geq \langle \widetilde{\boldsymbol{a}}_{k,i}, \boldsymbol{z}\rangle, \forall i \in \{1,\ldots,\widetilde{I}_k\} \setminus \{i_k\}, k = 1,\ldots,\widetilde{K}\right\}$$

that

$$\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) = \sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k)\langle \widetilde{\boldsymbol{a}}_{k,i_k}, \boldsymbol{z}\rangle.$$

But

$$\begin{aligned}
&\left\{\boldsymbol{z} \in \mathbb{R}^d_+ : \langle \widetilde{\boldsymbol{a}}_{k,i_k}, \boldsymbol{z}\rangle \geq \langle \widetilde{\boldsymbol{a}}_{k,i}, \boldsymbol{z}\rangle, \forall i \in \{1,\ldots,\widetilde{I}_k\} \setminus \{i_k\}, k = 1,\ldots,\widetilde{K}\right\}\\
&\quad = \left\{\boldsymbol{z} \in \mathbb{R}^d_+ : \langle \widetilde{\boldsymbol{a}}_{k,i_k} - \widetilde{\boldsymbol{a}}_{k,i}, \boldsymbol{z}\rangle \geq 0, \forall i \in \{1,\ldots,\widetilde{I}_k\} \setminus \{i_k\}, k = 1,\ldots,\widetilde{K}\right\}\\
&\quad = \operatorname{dual}\left(\left\{\widetilde{\boldsymbol{a}}_{k,i_k} - \widetilde{\boldsymbol{a}}_{k,i} : i \in \{1,\ldots,\widetilde{I}_k\} \setminus \{i_k\}, k = 1,\ldots,\widetilde{K}\right\}\right) \cap \mathbb{R}^d_+\\
&\quad = \operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+.
\end{aligned} \tag{B.8}$$

By (B.8), it is straightforward to verify that $\bigcup_{(i_k)\in\widetilde{\mathcal{I}}} \operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+ = \mathbb{R}^d_+$, and that if $(i_k) \in \widetilde{\mathcal{I}}$, $(i'_k) \in \widetilde{\mathcal{I}}$, $(i_k) \neq (i'_k)$, then $\operatorname{int}(\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) \cap \operatorname{int}(\operatorname{dual}(A_{(i'_k)}) \cap \mathbb{R}^d_+) = \emptyset$. We have thus expressed $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z})$ by a local representation as in Lemma 3.3(ii) over the partition $\{\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+ : (i_k) \in \widetilde{\mathcal{I}}\}$ of $\mathbb{R}^d_+$. The last part of statement (i) is due to Lemma 3.3(iii). We have thus proved statement (i).

By statement (i), $\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0$ is equivalent to $\sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k)\langle \widetilde{\boldsymbol{a}}_{k,i_k}, \boldsymbol{z}\rangle \geq 0$ for all $\boldsymbol{z} \in \operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+$. This is equivalent to

$$\begin{aligned}
\sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k)\widetilde{\boldsymbol{a}}_{k,i_k} \in \operatorname{dual}(\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) &= \operatorname{dual}(\operatorname{dual}(A_{(i_k)} + \mathbb{R}^d_+))\\
&= \operatorname{cone}(A_{(i_k)}) + \mathbb{R}^d_+.
\end{aligned}$$

Therefore,

$$\widetilde{s}_{\boldsymbol{y}}(\boldsymbol{z}) \geq 0\ \forall \boldsymbol{z} \in \operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+ \quad \Longleftrightarrow \quad \exists \boldsymbol{u} \in \operatorname{cone}(A_{(i_k)}) \text{ s.t. } \sum_{k=1}^{\widetilde{K}} (\langle \boldsymbol{y}, \widetilde{\boldsymbol{w}}_k\rangle + \widetilde{z}_k)\widetilde{\boldsymbol{a}}_{k,i_k} \geq \boldsymbol{u},$$

which is equivalent to the existence of $\eta^{(i_k)}_{\boldsymbol{v}} \geq 0$ for each $\boldsymbol{v} \in A_{(i_k)}$ such that (3.10) holds. We have now proved statement (ii).

To prove statement (iii), let us assume without loss of generality that $\boldsymbol{0} \notin A_{(i_k)}$ (otherwise it can be removed without changing $\operatorname{dual}(A_{(i_k)})$). Suppose first that $\operatorname{int}(\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) = \emptyset$. Since $\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+$ is a polyhedral cone, it is contained in a $(d-1)$-dimensional subspace, say $\{\boldsymbol{w} \in \mathbb{R}^n : \langle \boldsymbol{u}_0, \boldsymbol{w}\rangle = 0\}$ for some $\boldsymbol{u}_0 \neq \boldsymbol{0}$, which implies that both $\boldsymbol{u}_0$ and $-\boldsymbol{u}_0$ are elements of $\operatorname{dual}(\operatorname{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) = \operatorname{cone}(A_{(i_k)}) + \mathbb{R}^d_+$. This implies that there exist $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \operatorname{cone}(A_{(i_k)})$

such that $\boldsymbol{u}_1 \le \boldsymbol{u}_0, \boldsymbol{u}_2 \le -\boldsymbol{u}_0$. Thus, $\boldsymbol{u} := \boldsymbol{u}_1 + \boldsymbol{u}_2 \le \boldsymbol{0}$, and $\boldsymbol{u} \in \mathrm{cone}(A_{(i_k)})$. This implies that for each $\boldsymbol{v} \in A_{(i_k)}$ there exists $\eta_{\boldsymbol{v}}^{(i_k)} \ge 0$ such that $\boldsymbol{u} = \sum_{\boldsymbol{v}\in A_{(i_k)}} \eta_{\boldsymbol{v}}^{(i_k)}\boldsymbol{v}$, while $(\eta_{\boldsymbol{v}}^{(i_k)})_{\boldsymbol{v}\in A_{(i_k)}}$ are not all identically zero (otherwise we will have $\boldsymbol{u}_1 = \boldsymbol{u}_2 = \boldsymbol{0}$, and thus $\boldsymbol{u}_0 = \boldsymbol{0}$). We are allowed to scale $(\eta_{\boldsymbol{v}}^{(i_k)})_{\boldsymbol{v}\in A_{(i_k)}}$ by the same positive factor to make $\sum_{\boldsymbol{v}\in A_{(i_k)}} \eta_{\boldsymbol{v}}^{(i_k)} = 1$, thus condition (3.11) holds.

Conversely, suppose that the condition (3.11) holds, and let $(\eta_{\boldsymbol{v}}^{(i_k)})_{\boldsymbol{v}\in A_{(i_k)}}$ be given as in (3.11). Define $\boldsymbol{u} := \sum_{\boldsymbol{v}\in A_{(i_k)}} \eta_{\boldsymbol{v}}^{(i_k)}\boldsymbol{v}$. By (3.11), $\boldsymbol{u} \le \boldsymbol{0}$. If $\boldsymbol{u} \ne \boldsymbol{0}$, then $\boldsymbol{u} \in \mathrm{cone}(A_{(i_k)}) + \mathbb{R}^d_+$, $-\boldsymbol{u} \in \mathbb{R}^d_+ \subset \mathrm{cone}(A_{(i_k)}) + \mathbb{R}^d_+$, and $\mathrm{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+ = \mathrm{dual}(\mathrm{cone}(A_{(i_k)})) \cap \mathbb{R}^d_+ = \mathrm{dual}(\mathrm{cone}(A_{(i_k)}) + \mathbb{R}^d_+)$ is a subset of the subspace $\{\boldsymbol{w} \in \mathbb{R}^d : \langle \boldsymbol{u}, \boldsymbol{w}\rangle = 0\}$, thus $\mathrm{int}(\mathrm{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) = \emptyset$. If $\boldsymbol{u} = \boldsymbol{0}$, then let $\boldsymbol{v}' \in A_{(i_k)}$ such that $\eta_{\boldsymbol{v}'}^{(i_k)} > 0$. Thus, $\eta_{\boldsymbol{v}'}^{(i_k)}\boldsymbol{v}' \ne \boldsymbol{0}$ since $\boldsymbol{0} \notin A_{(i_k)}$, $\eta_{\boldsymbol{v}'}^{(i_k)}\boldsymbol{v}' \in \mathrm{cone}(A_{(i_k)}) + \mathbb{R}^d_+$, and

$$-\eta_{\boldsymbol{v}'}^{(i_k)}\boldsymbol{v}' = \boldsymbol{u} - \eta_{\boldsymbol{v}'}^{(i_k)}\boldsymbol{v}' = \sum_{\boldsymbol{v}\in A_{(i_k)}, \boldsymbol{v}\ne\boldsymbol{v}'} \eta_{\boldsymbol{v}}^{(i_k)}\boldsymbol{v} \in \mathrm{cone}(A_{(i_k)}) + \mathbb{R}^d_+.$$

Since both $\eta_{\boldsymbol{v}'}^{(i_k)}\boldsymbol{v}'$ and $-\eta_{\boldsymbol{v}'}^{(i_k)}\boldsymbol{v}'$ are elements of $\mathrm{cone}(A_{(i_k)}) + \mathbb{R}^d_+$, one can deduce that $\mathrm{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+ = \mathrm{dual}(\mathrm{cone}(A_{(i_k)})) \cap \mathbb{R}^d_+ = \mathrm{dual}(\mathrm{cone}(A_{(i_k)}) + \mathbb{R}^d_+)$ is a subset of the subspace $\{\boldsymbol{w} \in \mathbb{R}^d : \langle \boldsymbol{v}', \boldsymbol{w}\rangle = 0\}$, thus $\mathrm{int}(\mathrm{dual}(A_{(i_k)}) \cap \mathbb{R}^d_+) = \emptyset$. The proof is now complete. □

Proof of Lemma 3.12. Let p denote the objective function of problem (3.14). To prove Lemma 3.12, we show that each minimizer of problem (3.14) gives a feasible point of problem (3.13) with identical objective value and vice versa. Suppose that $(\boldsymbol{x}^\star, (\lambda_k^\star), (\zeta_k^\star), (\delta_{k,i}^\star), (\iota_{k,i}^\star))$ is a minimizer of problem (3.14) (which exists since (3.14) can be considered as the minimum of a finite collection of LP problems, each corresponding to a feasible combination of binary variables $(\iota_{k,i})$). By the feasibility conditions of (3.14), for each k such that $\xi_k = 1$, $\lambda_k^\star \ge \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\}$, and since $\lambda_k^\star$ is the minimum of such λ_k, $\lambda_k^\star = \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\}$. For each k such that $\xi_k = -1$, $\zeta_k^\star \ge \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\}$. By $\iota_{k,i} \in \{0, 1\}$ for all $1 \le i \le I_k$ and $\sum_{i=1}^{I_k} \iota_{k,i} = 1$, there exists a unique $i' \in \{1, \ldots, I_k\}$ such that $\iota_{k,i'} = 1$. Subsequently, $\delta_{k,i'} = 0$ and $\zeta_k^\star = \langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i'}$ and thus $\zeta_k^\star = \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\}$. We have

$$\begin{aligned}
p(\boldsymbol{x}^\star, (\lambda_k^\star), (\zeta_k^\star), (\delta_{k,i}^\star), (\iota_{k,i}^\star)) &= \sum_{k=1,\ldots,K, \xi_k=1} \lambda_k^\star + \sum_{k=1,\ldots,K, \xi_k=-1} -\zeta_k^\star \\
&= \sum_{k=1,\ldots,K, \xi_k=1} \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\} \\
&\quad + \sum_{k=1,\ldots,K, \xi_k=-1} - \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\} \\
&= \sum_{k=1,\ldots,K} \xi_k \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\} \\
&= h(\boldsymbol{x}^\star).
\end{aligned}$$

Hence, $\boldsymbol{x}^\star$ is a feasible point of problem (3.13) with identical objective value. Conversely, let $\boldsymbol{x}^\star$ be a minimizer of problem (3.13). For each k such that $\xi_k = 1$, define $\lambda_k^\star := \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\}$. For each k such that $\xi_k = -1$, define $\zeta_k^\star := \max_{1\le i\le I_k} \{\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle + b_{k,i}\}$ and for $i = 1, \ldots, I_k$, define $\delta_{k,i}^\star := \zeta_k^\star - \langle \boldsymbol{a}_{k,i}, \boldsymbol{x}^\star\rangle - b_{k,i}$. It is clear that $\delta_{k,i}^\star \ge 0$. $\delta_{k,i}^\star \le M_{k,i}$ is guaranteed by the definition of $M_{k,i}$ in (3.12). Let $i' \in \{1, \ldots, I_k\}$ such that $\delta_{k,i'}^\star = 0$, and define $\iota_{k,i'}^\star := 1$ and $\iota_{k,i}^\star := 0$ for all $i \ne i'$. It is straightforward to verify that $(\boldsymbol{x}^\star, (\lambda_k^\star), (\zeta_k^\star), (\delta_{k,i}^\star), (\iota_{k,i}^\star))$ is a feasible point of problem (3.14). Moreover, by our definitions

$$\begin{aligned}
h(\boldsymbol{x}^\star) &= \sum_{k=1,\ldots,K, \xi_k=1} \lambda_k^\star + \sum_{k=1,\ldots,K, \xi_k=-1} -\zeta_k^\star \\
&= p(\boldsymbol{x}^\star, (\lambda_k^\star), (\zeta_k^\star), (\delta_{k,i}^\star), (\iota_{k,i}^\star)).
\end{aligned}$$

Hence, $(\boldsymbol{x}^\star, (\lambda_k^\star), (\zeta_k^\star), (\delta_{k,i}^\star), (\iota_{k,i}^\star))$ is a feasible point of problem (3.14) with identical objective value. The proof is now complete. □

B.1.2. *Proof of Theorem 3.16.*

Proof of Theorem 3.16. By the assumption about $\underline{\phi}$ in Remark 3.14, there exist c_0 and $\boldsymbol{y}_0$ such that $c_0 + \langle \boldsymbol{y}_0, \boldsymbol{g}\rangle \geq -f$ and $\underline{\phi} = -c_0 - \pi(\boldsymbol{y}_0)$. For any c and $\boldsymbol{y}$ such that $c + \langle \boldsymbol{y}, \boldsymbol{g}\rangle \geq f$, it holds that

$$c_0 + c + \langle \boldsymbol{y}_0 + \boldsymbol{y}, \boldsymbol{g}\rangle \geq 0$$

and thus

$$\langle \boldsymbol{y}_0 + \boldsymbol{y}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_0 + \boldsymbol{y}) \geq -c_0 - c - \pi(\boldsymbol{y}_0 + \boldsymbol{y}).$$

By Assumption 2.1, $-c_0 - c - \pi(\boldsymbol{y}_0 + \boldsymbol{y}) \leq 0$, and so

$$-c_0 - c \leq \pi(\boldsymbol{y}_0 + \boldsymbol{y}) \leq \pi(\boldsymbol{y}_0) + \pi(\boldsymbol{y}),$$

thus $\underline{\phi} = -c_0 - \pi(\boldsymbol{y}_0) \leq c + \pi(\boldsymbol{y})$. This implies that $\underline{\phi} \leq \phi(f)$. Let $\binom{\widehat{c}}{\widehat{\boldsymbol{y}}} \in S^\star$. We have $\widehat{c} + \pi(\widehat{\boldsymbol{y}}) = \phi(f) > \underline{\phi} - \tau$. Let $\sigma^{(r)}$ denote the system of linear inequalities σ at iteration r. By letting $\widehat{\boldsymbol{y}}^+ = \max\{\widehat{\boldsymbol{y}}, \mathbf{0}\}$, $\widehat{\boldsymbol{y}}^- = \max\{-\widehat{\boldsymbol{y}}, \mathbf{0}\}$ where $\max\{\cdot\}$ is applied component-wise, $(\widehat{c}, \widehat{\boldsymbol{y}}^+, \widehat{\boldsymbol{y}}^-)$ satisfies all constraints in $\sigma^{(r)}$ for all r. We have

$$\begin{aligned}
\underline{\varphi}^{(r)} &= \inf_{(c,\boldsymbol{y}^+,\boldsymbol{y}^-) \text{ satisfies } \sigma^{(r)}} c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}}\rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}}\rangle \\
&= \inf_{(c,\boldsymbol{y}^+,\boldsymbol{y}^-) \text{ satisfies } \sigma^{(r)}} c + \pi(\boldsymbol{y}^+ - \boldsymbol{y}^-) \\
&\leq \widehat{c} + \pi(\widehat{\boldsymbol{y}}^+ - \widehat{\boldsymbol{y}}^-) \\
&= \widehat{c} + \pi(\widehat{\boldsymbol{y}}) \\
&= \phi(f).
\end{aligned} \tag{B.9}$$

$\underline{\varphi}^{(r)}$ is non-decreasing in r because more constraints are added to σ. Moreover, by Proposition 3.9(iv) and the assumption on $\overline{\boldsymbol{x}}$, for any $\boldsymbol{y}$ that satisfies $\widetilde{\sigma}$, it holds that $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x}) = \inf_{\mathbf{0}\leq\boldsymbol{x}\leq\overline{\boldsymbol{x}}} s_{\boldsymbol{y}}(\boldsymbol{x})$. By Definition 3.7, for all r and for any $\boldsymbol{x} \in \Omega$,

$$\begin{aligned}
& c^{(r)} - s^{(r)} + \langle \boldsymbol{y}^{(r)}, \boldsymbol{g}(\boldsymbol{x})\rangle - f(\boldsymbol{x}) \\
=& c^{(r)} - c^{(r)} - \inf_{\mathbf{0}\leq\boldsymbol{x}'\leq\overline{\boldsymbol{x}}} s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x}') + \langle \boldsymbol{y}^{(r)}, \boldsymbol{g}(\boldsymbol{x})\rangle - f(\boldsymbol{x}) \\
=& \langle \boldsymbol{y}^{(r)}, \boldsymbol{g}(\boldsymbol{x})\rangle - f(\boldsymbol{x}) - \inf_{\boldsymbol{x}'\in\mathbb{R}^d_+} s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x}') \\
=& \langle \boldsymbol{y}^{(r)}, \boldsymbol{g}(\boldsymbol{x})\rangle - f(\boldsymbol{x}) - \inf_{\boldsymbol{x}'\in\mathbb{R}^d_+} \Big\{\langle \boldsymbol{y}^{(r)}, \boldsymbol{g}(\boldsymbol{x}')\rangle - f(\boldsymbol{x}')\Big\} \\
\geq& 0,
\end{aligned} \tag{B.10}$$

and thus by Line 9, $\underline{\varphi}^{(r)} - s^{(r)} = c^{(r)} - s^{(r)} + \pi(\boldsymbol{y}^{(r)}) \geq \phi(f)$. This and (B.9) also show that $s^{(r)} \leq 0$. We have proved statement (i).

If Algorithm 1 terminates, then by (B.10), for any $\boldsymbol{x} \in \mathbb{R}^d_+$,

$$c^\star + \langle \boldsymbol{y}^\star, \boldsymbol{g}(\boldsymbol{x})\rangle - f(\boldsymbol{x}) = c^{(r-1)} - s^{(r-1)} + \langle \boldsymbol{y}^{(r-1)}, \boldsymbol{g}(\boldsymbol{x})\rangle - f(\boldsymbol{x}) \geq 0.$$

Therefore, $(c^\star, \boldsymbol{y}^\star)$ is feasible for (2.4) and $\phi(f)^{\mathrm{LB}} \leq \phi(f) \leq \phi(f)^{\mathrm{UB}}$ follows directly from statement (i). Since at termination, $s^{(r-1)} \geq -\varepsilon$, we have $\phi(f)^{\mathrm{UB}} - \phi(f)^{\mathrm{LB}} \leq \varepsilon$ and $(c^\star, \boldsymbol{y}^\star)$ is ε-optimal.

We now show that Algorithm 1 terminates. Same as in the proof of Proposition 3.9(iv), let us fix a $\boldsymbol{y}$ such that $\inf_{\boldsymbol{x}\in\mathbb{R}^d_+} s_{\boldsymbol{y}}(\boldsymbol{x}) > -\infty$ and express $s_{\boldsymbol{y}}(\boldsymbol{x})$ by (3.8). We restrict $s_{\boldsymbol{y}}(\boldsymbol{x})$ to those $\mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}$. Let $\mathcal{I} := \{(i_k)_{k=1:K} : 1 \leq i_k \leq I_k\}$. Same as in the proof of Lemma 3.3(ii), $s_{\boldsymbol{y}}(\boldsymbol{x})$ has a local representation as in (B.1), with ξ_k replaced by $\langle \boldsymbol{y}, \boldsymbol{w}_k\rangle + z_k$, over the finite partition $\mathfrak{C} := \Big\{\bigcap_{k=1}^K \Omega_{k,i_k} : (i_k) \in \mathcal{I}\Big\}$ of $\{\boldsymbol{x} \in \mathbb{R}^d_+ : \mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}\}$, where $\Omega_{k,i} := \{\mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}} :$

$\langle \boldsymbol{a}_{k,i}, \boldsymbol{x}\rangle + b_{k,i} \geq \langle \boldsymbol{a}_{k,i'}, \boldsymbol{x}\rangle + b_{k,i'}\ \forall 1 \leq i' \leq I_k\}$. Notice that the partition $\mathfrak{C}$ does not depend on the value of $\boldsymbol{y}$, and each $C \in \mathfrak{C}$ is a polytope. Let $\mathfrak{F} := \{F \neq \emptyset$ is a face of some $C \in \mathfrak{C}\}$. By Theorem 18.2 of Rockafellar [60],

$$\bigcup_{F\in\mathfrak{F}} \mathrm{relint}(F) = \bigcup_{C\in\mathfrak{C}} C = \{\boldsymbol{x} \in \mathbb{R}^d_+ : \mathbf{0} \leq \boldsymbol{x} \leq \overline{\boldsymbol{x}}\}. \tag{B.11}$$

By Theorem 19.1 of Rockafellar [60], $|\mathfrak{F}| < \infty$. Let $\boldsymbol{x}^{(r)}$ be a minimizer of the MILP problem in Line 12, that is, $c^{(r)} + s_{\boldsymbol{y}^{(r)}}(\boldsymbol{x}^{(r)}) = s^{(r)}$.

We prove that either Algorithm 1 terminates, or for each $F \in \mathfrak{F}$, there exists at most one $r \in \mathbb{N}$ such that $\boldsymbol{x}^{(r)} \in \mathrm{relint}(F)$. Suppose, for the sake of contradiction, that Algorithm 1 does not terminate, and that there exists $r, l \in \mathbb{N}, r < l$, and $\boldsymbol{x}^{(r)}, \boldsymbol{x}^{(l)} \in \mathrm{relint}(F)$ for some $F \in \mathfrak{F}$, which is a face of some $C \in \mathfrak{C}$. Since $\boldsymbol{x}^{(r)} \in X^{(r)}$, we have

$$c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\boldsymbol{x}^{(r)}) \geq 0 \tag{B.12}$$

by Line 8. We also have

$$c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\boldsymbol{x}^{(l)}) = s^{(l)} < 0, \tag{B.13}$$

since otherwise Algorithm 1 will terminate at the l-th iteration. For every $\lambda \in \mathbb{R}$, let $\boldsymbol{x}_\lambda := (1 - \lambda)\boldsymbol{x}^{(r)} + \lambda \boldsymbol{x}^{(l)}$. Since $\boldsymbol{x}^{(r)}, \boldsymbol{x}^{(l)} \in \mathrm{relint}(F)$, there exists $\widehat{\lambda} > 1$ such that $\boldsymbol{x}_{\widehat{\lambda}} \in F \subset C$. By the local representation (B.1), $c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\cdot)$ is an affine function when restricted to the set C. Therefore, by (B.12) and (B.13), we have

$$\begin{aligned} c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\boldsymbol{x}_{\widehat{\lambda}}) =& (1-\widehat{\lambda})\Big(c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\boldsymbol{x}^{(r)})\Big) + \widehat{\lambda}\Big(c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\boldsymbol{x}^{(l)})\Big) \\ \leq& \widehat{\lambda}\Big(c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\boldsymbol{x}^{(l)})\Big) \\ <& c^{(l)} + s_{\boldsymbol{y}^{(l)}}(\boldsymbol{x}^{(l)}), \end{aligned}$$

contradicting the fact that $\boldsymbol{x}^{(l)}$ is a minimizer of the MILP problem in Line 12.

Since for each r, $\boldsymbol{x}^{(r)} \in \mathrm{relint}(F)$ for some $F \in \mathfrak{F}$ as a consequence of (B.11), and since $|\mathfrak{F}| < \infty$, Algorithm 1 terminates eventually. The proof of statement (ii) is now complete.

Finally, if Line 17 of Algorithm 1 is reached, then $\underline{\phi} > \underline{\varphi}^{(r-1)} - s^{(r-1)}$. By (B.10),

$$c^{(r-1)} - s^{(r-1)} + \langle \boldsymbol{y}^{(r-1)}, \boldsymbol{g}\rangle \geq f.$$

By the assumption about $\underline{\phi}$ in Remark 3.14, there exist c_0 and $\boldsymbol{y}_0$ such that $c_0 + \langle \boldsymbol{y}_0, \boldsymbol{g}\rangle \geq -f$ and $\underline{\phi} = -c_0 - \pi(\boldsymbol{y}_0)$. Hence, we have

$$c_0 + c^{(r-1)} - s^{(r-1)} + \langle \boldsymbol{y}_0 + \boldsymbol{y}^{(r-1)}, \boldsymbol{g}\rangle \geq 0,$$

and thus

$$\begin{aligned} \langle \boldsymbol{y}_0 + \boldsymbol{y}^{(r-1)}, \boldsymbol{g}\rangle - \pi(\boldsymbol{y}_0 + \boldsymbol{y}^{(r-1)}) \geq& - c_0 - c^{(r-1)} + s^{(r-1)} - \pi(\boldsymbol{y}_0 + \boldsymbol{y}^{(r-1)}) \\ \geq& - c_0 - \pi(\boldsymbol{y}_0) - c^{(r-1)} + s^{(r-1)} - \pi(\boldsymbol{y}^{(r-1)}) \\ =& \underline{\phi} - \underline{\varphi}^{(r-1)} + s^{(r-1)} > 0, \end{aligned}$$

which is a violation of Assumption 2.1. The proof is now complete. □

B.2. **Proof of results in Setting 2.**

Proof of Theorem 3.24. If Assumption 2.1 holds, then by the same argument as in the proof of Theorem 3.16(i), $\underline{\phi} \leq \phi(f)$. Hence, $\underline{\varphi}^{(0)} \leq \phi(f) \leq \overline{\varphi}^{(0)}$ and $c^{\star(0)} + \langle \boldsymbol{y}^{\star(0)}, \boldsymbol{g}\rangle \geq f$ follow from assumptions.

For $r \geq 1$, suppose that $\underline{\varphi}^{(r-1)} \leq \phi(f)$. Then, $\underline{\varphi}^{(r)} \neq \underline{\varphi}^{(r-1)}$ only when Line 13 is reached. This implies that $\sigma(\overline{c}, \overline{\boldsymbol{y}}, \underline{\varphi}^{(r-1)}, \varphi^{(r)}, X) = \emptyset$ by Lemma 3.20. By the assumption that

$$\{\left(\begin{smallmatrix} c \\ \boldsymbol{y} \end{smallmatrix}\right) : |c| \leq \overline{c}, -\overline{\boldsymbol{y}} \leq \boldsymbol{y} \leq \overline{\boldsymbol{y}}\} \cap S^\star \neq \emptyset$$

in Remark 3.22, there exists $\binom{\widehat{c}}{\widehat{\boldsymbol{y}}} \in S^\star$ that satisfies $|\widehat{c}| \le \overline{c}$, $\mathbf{0} \le \widehat{\boldsymbol{y}}^+ \le \overline{\boldsymbol{y}}$, $\mathbf{0} \le \widehat{\boldsymbol{y}}^- \le \overline{\boldsymbol{y}}$, where $\widehat{\boldsymbol{y}}^+ := \max\{\widehat{\boldsymbol{y}}, \mathbf{0}\}$, $\widehat{\boldsymbol{y}}^- := \max\{-\widehat{\boldsymbol{y}}, \mathbf{0}\}$ and $\max\{\cdot\}$ is applied component-wise. Moreover, $\widehat{c} + \langle \widehat{\boldsymbol{y}}^+ - \widehat{\boldsymbol{y}}^-, \boldsymbol{g}(\boldsymbol{x})\rangle \ge f(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$ and

$$\widehat{c} + \langle \widehat{\boldsymbol{y}}^+, \overline{\boldsymbol{\pi}}\rangle - \langle \widehat{\boldsymbol{y}}^-, \underline{\boldsymbol{\pi}}\rangle = \widehat{c} + \pi(\widehat{\boldsymbol{y}}) = \phi(f). \tag{B.14}$$

By (B.14) and by the assumption that $\underline{\varphi}^{(r-1)} \le \phi(f)$, we have that $\sigma(\overline{c}, \overline{\boldsymbol{y}}, \underline{\varphi}^{(r-1)}, \varphi^{(r)}, X) = \emptyset$ implies $\varphi^{(r)} < \phi(f)$. Therefore, by Line 13, $\underline{\varphi}^{(r)} > \varphi^{(r)} > \underline{\varphi}^{(r-1)}$. Since $(\widehat{c}, \widehat{\boldsymbol{y}}^+, \widehat{\boldsymbol{y}}^-)$ also satisfies all constraints in the LP problem in Line 13, we have, again by (B.14), that $\underline{\varphi}^{(r)} \le \phi(f)$. By induction, we have proved that $\underline{\varphi}^{(r)}$ is non-decreasing in r and $\underline{\varphi}^{(r)} \le \phi(f)$ for all r.

For $r \ge 1$, $\overline{\varphi}^{(r)} \ne \overline{\varphi}^{(r-1)}$, $c^{\star(r)} \ne c^{\star(r-1)}$, or $\boldsymbol{y}^{\star(r)} \ne \boldsymbol{y}^{\star(r-1)}$ only if Line 23 is reached. By Line 22 and Line 23, $\overline{\varphi}^{(r)} < \overline{\varphi}^{(r-1)}$. By the same reasoning as in the proof of Theorem 3.16(i) in equation (B.10), we have $c^{\star(r)} + \langle \boldsymbol{y}^{\star(r)}, \boldsymbol{g}\rangle \ge f$ and $\overline{\varphi}^{(r)} \ge \phi(f)$. We have thus proved statement (i).

If Assumption 2.1 holds and Algorithm 2 terminates, then $\phi(f)^{\mathrm{LB}} \le \phi(f) \le \phi(f)^{\mathrm{UB}}$ and the feasibility and ε-optimality of $(c^\star, \boldsymbol{y}^\star)$ follow directly from statement (i) and Line 5. Thus, we only need to show that Algorithm 2 terminates. Notice that the so-called Strong Slater Condition in Theorem 1 of Betrò [11] holds because for some $\eta > 0$, one can choose any $\binom{\widehat{c}}{\widehat{\boldsymbol{y}}} \in S$ and $(\widehat{c} + \eta) + \langle \widehat{\boldsymbol{y}}, \boldsymbol{g}\rangle \ge f + \eta$. Moreover, under Assumption 3.18, $\sup_{\boldsymbol{x}\in\Omega} \|\boldsymbol{g}(\boldsymbol{x})\| < \infty$. Suppose, for the sake of contradiction, that Algorithm 2 loops infinitely and does not terminate. Then, one can deduce that after finitely many iterations, Line 13 is never reached, since each time Line 13 is reached, $\overline{\varphi}^{(r)} - \underline{\varphi}^{(r)} \le \frac{3}{4}(\overline{\varphi}^{(r-1)} - \underline{\varphi}^{(r-1)})$ by Line 7 and Line 9. Similarly, Line 23 is never reached again after finitely many iterations since each time Line 23 is reached $\overline{\varphi}^{(r)} - \underline{\varphi}^{(r)} \le \overline{\varphi}^{(r-1)} - \underline{\varphi}^{(r-1)} - \varepsilon$. The rest of the proof follows exactly as the proof of Theorem 1 in Betrò [11]. The proof of statement (ii) is now complete.

For statement (iii), notice that since $\underline{\varphi}^{(0)} = \underline{\phi} - \tau \le \phi(f) - \tau < \phi(f)^{\mathrm{LB}}$, Line 13 is reached at least once before termination. Thus, $c^\dagger$ and $\boldsymbol{y}^\dagger$ are defined. Let $\boldsymbol{y}^{+\dagger} := \max\{\boldsymbol{y}^\dagger, \mathbf{0}\}$, $\boldsymbol{y}^{-\dagger} := \max\{-\boldsymbol{y}^\dagger, \mathbf{0}\}$. Then, by Line 13, $(c^\dagger, \boldsymbol{y}^{+\dagger}, \boldsymbol{y}^{-\dagger})$ is an optimal solution of the LP problem:

$$\begin{aligned} \text{minimize} \quad & c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}}\rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}}\rangle \\ \text{subject to} \quad & c + \langle \boldsymbol{y}^+ - \boldsymbol{y}^-, \boldsymbol{g}(\boldsymbol{x})\rangle \ge f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in X^\dagger, \\ & -\overline{c} \le c \le \overline{c}, \mathbf{0} \le \boldsymbol{y}^+ \le \overline{\boldsymbol{y}}, \mathbf{0} \le \boldsymbol{y}^- \le \overline{\boldsymbol{y}}. \end{aligned} \tag{B.15}$$

Thus, since $\underline{\varphi}^{(r)}$ is updated whenever $(c^\dagger, \boldsymbol{y}^\dagger)$ are updated, we have $\phi(f)^{\mathrm{LB}} = c^\dagger + \langle \boldsymbol{y}^{+\dagger}, \overline{\boldsymbol{\pi}}\rangle - \langle \boldsymbol{y}^{-\dagger}, \underline{\boldsymbol{\pi}}\rangle$, and $c^\dagger + \langle \boldsymbol{y}^{+\dagger} - \boldsymbol{y}^{-\dagger}, \boldsymbol{g}(\boldsymbol{x})\rangle \ge f(\boldsymbol{x})$, $\forall \boldsymbol{x} \in X^\dagger$. Let

$$\widetilde{B} := \left\{ \begin{pmatrix} c \\ \boldsymbol{y}^+ \\ \boldsymbol{y}^- \end{pmatrix} : -\overline{c} \le c \le \overline{c}, \mathbf{0} \le \boldsymbol{y}^+ \le \overline{\boldsymbol{y}}, \mathbf{0} \le \boldsymbol{y}^- \le \overline{\boldsymbol{y}} \right\} \subset \mathbb{R}^{2m+1}.$$

By the assumption of statement (iii), $-\overline{c} < c^\dagger < \overline{c}$, $\mathbf{0} \le \boldsymbol{y}^{+\dagger} < \overline{\boldsymbol{y}}$, $\mathbf{0} \le \boldsymbol{y}^{-\dagger} < \overline{\boldsymbol{y}}$, and we claim that $(c^\dagger, \boldsymbol{y}^{+\dagger}, \boldsymbol{y}^{-\dagger})$ is also optimal for the following LP problem:

$$\begin{aligned} \text{minimize} \quad & c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}}\rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}}\rangle \\ \text{subject to} \quad & c + \langle \boldsymbol{y}^+ - \boldsymbol{y}^-, \boldsymbol{g}(\boldsymbol{x})\rangle \ge f(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in X^\dagger, \\ & \boldsymbol{y}^+ \ge \mathbf{0}, \boldsymbol{y}^- \ge \mathbf{0}. \end{aligned} \tag{B.16}$$

Suppose, for the sake of contradiction, that (B.16) has optimal solution $(\widetilde{c}, \widetilde{\boldsymbol{y}}^+, \widetilde{\boldsymbol{y}}^-)$ with $\kappa := \widetilde{c} + \langle \widetilde{\boldsymbol{y}}^+, \overline{\boldsymbol{\pi}}\rangle - \langle \widetilde{\boldsymbol{y}}^-, \underline{\boldsymbol{\pi}}\rangle < \phi(f)^{\mathrm{LB}}$. Then, since $\phi(f)^{\mathrm{LB}}$ is the optimal value of (B.15), we have $\begin{pmatrix} \widetilde{c} \\ \widetilde{\boldsymbol{y}}^+ \\ \widetilde{\boldsymbol{y}}^- \end{pmatrix} \notin \widetilde{B}$, $\widetilde{\boldsymbol{y}}^+ \ge \mathbf{0}$, $\widetilde{\boldsymbol{y}}^- \ge \mathbf{0}$. Let $c_\lambda := \lambda c^\dagger + (1-\lambda)\widetilde{c}$, $\boldsymbol{y}_\lambda^+ := \lambda \boldsymbol{y}^{+\dagger} + (1-\lambda)\widetilde{\boldsymbol{y}}^+$, $\boldsymbol{y}_\lambda^- := \lambda \boldsymbol{y}^{-\dagger} + (1-\lambda)\widetilde{\boldsymbol{y}}^-$. Then, there exists some $\lambda \in (0,1)$, such that

$$\begin{pmatrix} c_\lambda \\ \boldsymbol{y}_\lambda^+ \\ \boldsymbol{y}_\lambda^- \end{pmatrix} = \lambda \begin{pmatrix} c^\dagger \\ \boldsymbol{y}^{+\dagger} \\ \boldsymbol{y}^{-\dagger} \end{pmatrix} + (1-\lambda) \begin{pmatrix} \widetilde{c} \\ \widetilde{\boldsymbol{y}}^+ \\ \widetilde{\boldsymbol{y}}^- \end{pmatrix} \in \widetilde{B},$$

$c_\lambda + \langle \boldsymbol{y}_\lambda^+ - \boldsymbol{y}_\lambda^-, \boldsymbol{g}(\boldsymbol{x})\rangle \geq f(\boldsymbol{x})$, $\forall \boldsymbol{x} \in X^\dagger$, and $c_\lambda + \langle \boldsymbol{y}_\lambda^+, \overline{\boldsymbol{\pi}}\rangle - \langle \boldsymbol{y}_\lambda^-, \underline{\boldsymbol{\pi}}\rangle = \lambda\phi(f)^{\mathrm{LB}} + (1-\lambda)\kappa < \phi(f)^{\mathrm{LB}}$, contradicting the optimality of $(c^\dagger, \boldsymbol{y}^{+\dagger}, \boldsymbol{y}^{-\dagger})$ for (B.15). Therefore, $(c^\dagger, \boldsymbol{y}^{+\dagger}, \boldsymbol{y}^{-\dagger})$ is also optimal for (B.16), whose corresponding dual LP problem is exactly (3.18). Then, an optimal solution $(\mu^\star_{\boldsymbol{x}})_{\boldsymbol{x}\in X^\dagger}$ of (3.18) exists, its corresponding finitely supported measure $\mu^\star$ is a probability measure, and satisfies $\underline{\pi}_j \leq \int_\Omega g_j \,\mathrm{d}\mu^\star \leq \overline{\pi}_j$ for $j = 1,\ldots,m$, and thus $\mu^\star \in \mathcal{Q}$. Moreover, due to strong duality of LP problems, $\int_\Omega f \,\mathrm{d}\mu^\star = \phi(f)^{\mathrm{LB}} \geq \phi(f) - \varepsilon$ by statement (ii), and $\mu^\star$ is ε-optimal for the right-hand side of (2.8) by Theorem 2.4(iii). We have completed the proof of statement (iii).

The proof of statement (iv) is exactly the same as the proof of Theorem 3.16(iii). The proof is now complete. □

Proof of Corollary 3.25. The proof that Algorithm 1 terminates is identical to the proof of Algorithm 3.16(ii). Hence, as in Theorem 3.16(ii), we have $\phi(f)^{\mathrm{LB}} \geq \phi(f) - \varepsilon$. Since Line 1 and Line 2 are not used, and that $\phi(f)^{\mathrm{LB}} \geq \underline{\phi} - \varepsilon > \underline{\phi} - \tau$, we have that $\phi(f)^{\mathrm{LB}}$ is the optimal value of the following LP problem:

$$\begin{aligned} \text{minimize} \quad & c + \langle \boldsymbol{y}^+, \overline{\boldsymbol{\pi}}\rangle - \langle \boldsymbol{y}^-, \underline{\boldsymbol{\pi}}\rangle \\ \text{subject to} \quad & c + \langle \boldsymbol{y}^+ - \boldsymbol{y}^-, \boldsymbol{g}(\boldsymbol{x})\rangle \geq f(\boldsymbol{x}) \quad \forall \boldsymbol{x} \in X = \bigcup_{l=0}^{r-1} X^{(l)}, \end{aligned}$$

whose dual LP problem is exactly (3.19). Consequently, by the same reasoning as in the proof of Theorem 3.24(iii), $\int_\Omega f \,\mathrm{d}\mu^\star = \phi(f)^{\mathrm{LB}} \geq \phi(f) - \varepsilon$, and $\mu^\star$ is ε-optimal for the right-hand side of (2.8) by Theorem 2.4(iii). □

REFERENCES

[1] B. Acciaio, M. Beiglböck, F. Penkner, and W. Schachermayer. A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Mathematical Finance*, 26(2):233–251, 2016. 1

[2] L. D. G. Aquino and C. Bernard. Bounds on multi-asset derivatives via neural networks. *Preprint, arXiv:1911.05523*, 2019. 2, 19

[3] A. Auslender, M. A. Goberna, and M. A. López. Penalty and smoothing methods for convex semi-infinite programming. *Mathematics of Operations Research*, 34(2):303–319, 2009. 19

[4] D. Bartl, M. Kupper, T. Lux, A. Papapantoleon, and S. Eckstein. Marginal and dependence uncertainty: bounds, optimal transport, and sharpness. *Preprint, arXiv:1709.00641*, 2017. 2, 4, 5

[5] D. Bartl, M. Kupper, D. J. Prömel, and L. Tangpi. Duality for pathwise superhedging in continuous time. *Finance and Stochastics*, 23(3):697–728, 2019. 1

[6] D. Bartl, M. Kupper, and A. Neufeld. Pathwise superhedging on prediction sets. *Finance and Stochastics*, 24(1):215–248, 2020. 1

[7] E. Bayraktar, Y.-J. Huang, and Z. Zhou. On hedging american options under model uncertainty. *SIAM Journal on Financial Mathematics*, 6(1):425–447, 2015. 1

[8] P. Beissner. Equilibrium prices and trade under ambiguous volatility. *Economic Theory*, 64(2): 213–238, 2017. 1

[9] P. Beissner and F. Riedel. Knight–Walras equilibria. *Preprint, arXiv:1605.04385*, 2016. 1

[10] P. Beissner and F. Riedel. Equilibria under Knightian price uncertainty. *Econometrica*, 87(1): 37–64, 2019. 1

[11] B. Betrò. An accelerated central cutting plane algorithm for linear semi-infinite programming. *Mathematical Programming*, 101(3):479–495, 2004. 7, 15, 17, 40

[12] J. M. Borwein and A. S. Lewis. Convergence of best entropy estimates. *SIAM Journal on Optimization*, 1(2):191–205, 1991. 19

[13] J. M. Borwein and A. S. Lewis. Duality relationships for entropy-like minimization problems. *SIAM Journal on Control and Optimization*, 29(2):325–338, 1991. 19

[14] J. M. Borwein and A. S. Lewis. Partially finite convex programming, part i: Quasi relative interiors and duality theory. *Mathematical Programming*, 57(1-3):15–48, 1992. 19, 30, 32

[15] B. Bouchard and M. Nutz. Arbitrage and duality in nondominated discrete-time models. *The Annals of Applied Probability*, 25(2):823–859, 2015. 1, 3, 4, 28, 29, 30
[16] B. Bouchard, S. Deng, and X. Tan. Superreplication with proportional transaction cost under model uncertainty. *Mathematical Finance*, 29(3):837–860, 2019. 1
[17] M. Burzoni, M. Frittelli, and M. Maggis. Universal arbitrage aggregator in discrete-time markets under uncertainty. *Finance and Stochastics*, 20(1):1–50, 2016. 1
[18] M. Burzoni, F. Riedel, and H. M. Soner. Viability and arbitrage under Knightian uncertainty. *Preprint, arXiv:1707.03335*, 2017. 1
[19] M. Burzoni, M. Frittelli, Z. Hou, M. Maggis, and J. Obłój. Pointwise arbitrage pricing theory in discrete time. *Mathematics of Operations Research*, 44(3):1034–1057, 2019. 1
[20] X. Chen, G. Deelstra, J. Dhaene, and M. Vanmaele. Static super-replicating strategies for a class of exotic options. *Insurance Math. Econom.*, 42:1067–1085, 2008. 2
[21] P. Cheridito, M. Kupper, and L. Tangpi. Duality formulas for robust pricing and hedging in discrete time. *SIAM Journal on Financial Mathematics*, 8(1):738–765, 2017. 1
[22] H. Cho, K.-K. Kim, and K. Lee. Computing lower bounds on basket option prices by discretizing semi-infinite linear programming. *Optimization Letters*, 10:1629–1644, 2016. 6
[23] D. Ciripoi, A. Löhne, and B. Weißing. Bensolve tools, version 1.3,(2019). gnu octave/matlab toolbox for calculus of convex polyhedra, calculus of polyhedral convex functions, global optimization, vector linear programming. 11
[24] R.-A. Dana and F. Riedel. Intertemporal equilibria with Knightian uncertainty. *Journal of Economic Theory*, 148(4):1582–1605, 2013. 1
[25] A. d'Aspremont and L. El Ghaoui. Static arbitrage bounds on basket option prices. *Math. Program.*, 106(3, Ser. A):467–489, 2006. 2, 5, 6
[26] S. Daum and R. Werner. A novel feasible discretization method for linear semi-infinite programming applied to basket option pricing. *Optimization*, 60(10-11):1379–1398, 2011. 2
[27] C. Dellacherie and P.-A. Meyer. Probabilities and potential. b, volume 72 of. *North-Holland Mathematics Studies*, 1982. 30
[28] J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance Math. Econom.*, 31:3–33, 2002. 2
[29] J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: applications. *Insurance Math. Econom.*, 31:133–161, 2002. 2
[30] Y. Dolinsky and A. Neufeld. Super-replication in fully incomplete markets. *Mathematical Finance*, 28(2):483–515, 2018. 1
[31] Y. Dolinsky and H. M. Soner. Martingale optimal transport and robust hedging in continuous time. *Probability Theory and Related Fields*, 160(1-2):391–427, 2014. 1
[32] S. Eckstein and M. Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Applied Mathematics & Optimization*, pages 1–29, 2019. 2, 19
[33] S. Eckstein, M. Kupper, and M. Pohl. Robust risk aggregation with neural networks. *Mathematical Finance*, 2018. 19
[34] S. Eckstein, G. Guo, T. Lim, and J. Obłój. Robust pricing and hedging of options on multiple assets and its numerics. *Preprint, arXiv:1909.03870*, 2019. 2, 19
[35] L. G. Epstein and S. Ji. Ambiguous volatility and asset pricing in continuous time. *The Review of Financial Studies*, 26(7):1740–1786, 2013. 1
[36] A. Galichon, P. Henry-Labordere, and N. Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. *The Annals of Applied Probability*, 24(1):312–336, 2014. 1
[37] M. A. Goberna and M. A. López. *Linear semi-infinite optimization*. John Wiley & Sons, 1998. 7, 14, 15
[38] M. A. Goberna and M. A. López. Recent contributions to linear semi-infinite optimization: an update. *Annals of Operations Research*, 271(1):237–278, 2018. 7
[39] L. Gurobi Optimization. Gurobi optimizer reference manual, 2020. URL http://www.gurobi.com. 12

[40] D. Hobson, P. Laurence, and T.-H. Wang. Static-arbitrage upper bounds for the prices of basket options. *Quant. Finance*, 5:329–342, 2005. 2
[41] D. Hobson, P. Laurence, and T.-H. Wang. Static-arbitrage optimal subreplicating strategies for basket options. *Insurance Math. Econom.*, 37:553–572, 2005. 2
[42] D. G. Hobson. Robust hedging of the lookback option. *Finance and Stochastics*, 2(4):329–347, 1998. 1
[43] J. Hu, J. Li, and S. Mehrotra. A data-driven functionally robust approach for simultaneous pricing and order quantity decisions with unknown demand function. *Operations Research*, 67 (6):1564–1585, 2019. 1
[44] N. Kahalé. Superreplication of financial derivatives via convex programming. *Management Science*, 63(7):2323–2339, 2017. 2
[45] H. G. Kellerer. Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 67(4):399–432, 1984. 5
[46] H. Le Verge. A note on chernikova's algorithm. 1992. 11
[47] C.-J. Lin, S.-C. Fang, and S.-Y. Wu. An unconstrained convex programming approach to linear semi-infinite programming. *SIAM Journal on Optimization*, 8(2):443–456, 1998. 19
[48] E. Lütkebohmert and J. Sester. Tightening robust price bounds for exotic derivatives. *Quantitative Finance*, 19(11):1797–1815, 2019. 1
[49] T. Lux and A. Papapantoleon. Improved Fréchet–Hoeffding bounds on d-copulas and applications in model-free finance. *Ann. Appl. Probab.*, 27:3633–3671, 2017. 2
[50] A. Neufeld and M. Nutz. Superreplication under volatility uncertainty for measurable claims. *Electronic journal of probability*, 18, 2013. 1
[51] J. Peña, X. Saynac, J. C. Vera, and L. F. Zuluaga. Computing general static-arbitrage bounds for european basket options via dantzig-wolfe decomposition. *Algorithmic Operations Research*, 5 (2):65–74, 2010. 2, 6
[52] J. Peña, J. C. Vera, and L. F. Zuluaga. Static-arbitrage lower bounds on the prices of basket options via linear programming. *Quant. Finance*, 10:819–827, 2010. 2, 5, 6
[53] D. Possamaï, G. Royer, and N. Touzi. On the robust superhedging of measurable claims. *Electronic Communications in Probability*, 18, 2013. 1
[54] G. Puccetti, L. Rüschendorf, and D. Manko. VaR bounds for joint portfolios with dependence constraints. *Depend. Model.*, 4:368–381, 2016. 2
[55] S. T. Rachev and L. Rüschendorf. Solution of some transportation problems with relaxed or additional constraints. *SIAM Journal on Control and Optimization*, 32:673–689, 1994. 5
[56] R. Reemtsen and J.-J. Rückmann. *Semi-infinite programming*, volume 25. Springer Science & Business Media, 1998. 7
[57] F. Riedel. Financial economics without probabilistic prior assumptions. *Decisions in Economics and Finance*, 38(1):75–91, 2015. 1
[58] L. Rigotti and C. Shannon. Uncertainty and risk in financial markets. *Econometrica*, 73(1): 203–243, 2005. 1
[59] R. T. Rockafellar. Integrals which are convex functionals. *Pacific journal of mathematics*, 24 (3):525–539, 1968. 19
[60] R. T. Rockafellar. *Convex analysis*. Number 28. Princeton university press, 1970. 4, 29, 34, 39
[61] P. Tankov. Improved Fréchet bounds and model-free pricing of multi-asset options. *J. Appl. Probab.*, 48:389–403, 2011. 2
[62] C. Villani. *Optimal Transport: Old and New*. Springer-Verlag, Berlin, 2009. 5
[63] Z. Yan, C. Cheng, K. Natarajan, and C.-P. Teo. "Marginal estimation+ price optimization" for multi-product pricing problems. Technical report, Working paper, 2018. 1
[64] D. A. Zaev. On the Monge-Kantorovich problem with additional linear constraints. *Matematicheskie Zametki*, 98:664–683, 2015. 5
[65] J. Zhen and D. Den Hertog. Computing the maximum volume inscribed ellipsoid of a polytopic projection. *INFORMS Journal on Computing*, 30(1):31–42, 2018. 15

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY, 21 NANYANG LINK, 637371 SINGAPORE
E-mail address: ariel.neufeld@ntu.edu.sg

DEPARTMENT OF MATHEMATICS, NATIONAL TECHNICAL UNIVERSITY OF ATHENS, ZOGRAFOU CAMPUS, 15780 ATHENS, GREECE & INSTITUTE OF APPLIED AND COMPUTATIONAL MATHEMATICS, FORTH, VASSILIKA VOUTON, 70013 HERAKLION, GREECE
E-mail address: papapan@math.ntua.gr

DIVISION OF MATHEMATICAL SCIENCES, NANYANG TECHNOLOGICAL UNIVERSITY, 21 NANYANG LINK, 637371 SINGAPORE
E-mail address: qikun001@e.ntu.edu.sg