Strong Duality in Risk-Constrained Nonconvex Functional Programming

Dionysis Kalogerias[†] and Spyridon Pougkakiotis[⋆]

[†]Department of EE—Yale University, USA.
[⋆]School of Science and Engineering—University of Dundee, UK.

E-mail(s): dionysis.kalogerias@yale.edu; spougkakiotis001@dundee.ac.uk;

Abstract

We show that risk-constrained functional optimization problems with general integrable nonconvex instantaneous reward/constraint functions exhibit strong duality, regardless of nonconvexity. We consider risk constraints featuring convex and positively homogeneous risk measures admitting dual representations with bounded risk envelopes, generalizing expectations. Popular risk measures supported within our setting include the conditional value-at-risk (CVaR), the mean-absolute deviation (MAD, including the non-monotone case), certain distributionally robust representations and more generally all real-valued coherent risk measures on the space $\boldsymbol{\mathcal{L}_1}$. We highlight the usefulness of our results by further discussing various generalizations of our base model, extensions for risk measures supported on $\boldsymbol{\mathcal{L}_{p>1}}$, implications in the context of mean-risk tradeoff models, as well as more specific applications in wireless systems resource allocation, and supervised constrained learning. Our core proof technique appears to be new and relies on risk conjugate duality in tandem with J. J. Uhl's weak extension of A. A. Lyapunov's convexity theorem for vector measures taking values in general infinite-dimensional Banach spaces.

Keywords: Lagrangian Duality, Strong Duality, Risk-Constrained Functional Programming, Nonconvex Optimization, Risk-Averse Optimization, Resource Allocation, Constrained Learning.

1 Introduction and Problem Setting

On some arbitrary base probability space $(\Omega, \mathscr{F}, \mu)$, consider a random element $\boldsymbol{H} : \Omega \to \mathcal{H} \triangleq \mathbb{R}^{N_{\boldsymbol{H}}}$ with induced Borel measure $\mathrm{P} : \mathscr{B}(\mathcal{H}) \to [0,1]$, modeling some observable random phenomenon, which we would like to optimally handle in a certain sense by making appropriate decisions. In particular, we are interested in risk-constrained nonconvex functional programs formulated as

$$\boxed{\begin{array}{rll} \infty > \mathsf{P}^* = & \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ & \text{subject to} & \boldsymbol{x} \le -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) \\ & & \boldsymbol{g}(\boldsymbol{x}) \ge \boldsymbol{0} \\ & & (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array}}, \tag{RCP}$$

where, with $\mathbf{C} \triangleq \mathbb{R}^N$, $g^o : \mathbf{C} \to \mathbb{R}$ and $\boldsymbol{g} : \mathbf{C} \to \mathbb{R}^{N_{\boldsymbol{g}}}$ are concave utility functions, $\boldsymbol{p} : \mathcal{H} \to \mathbf{R} \triangleq \mathbb{R}^{N_{\boldsymbol{p}}}$ is the allocation policy on observables $\boldsymbol{H}$, $\boldsymbol{f} : \mathbf{R} \times \mathcal{H} \to \mathbf{C}$ is a generally nonconvex *instantaneous* performance level score, measuring the quality of a policy $\boldsymbol{p}$ at each realization $\boldsymbol{H}$ in $\mathcal{H}$ and such that

---

This work has been supported by a Microsoft gift, and by the US National Science Foundation (NSF) under Grant CCF 2242215.

This work was developed while S. Pougkakiotis was a postdoctoral researcher with the Department of Electrical Engineering (EE), Yale University.

[†]Corresponding author.

$\boldsymbol{f}(\boldsymbol{p}(\cdot),\cdot) \in \mathcal{L}_1(\mathrm{P},\mathbf{C})$ on Π, and where $\boldsymbol{\rho} : \mathcal{L}_1(\mathrm{P},\mathbf{C}) \to \mathbf{C}$ is a finite-valued vector risk measure, which we assume that is *convex*[1]*, lower semicontinuous and positively homogeneous* in every dimension (i.e., component-wise), with the standardized convention that, for each $i \in \mathbb{N}_N^+$,

$$\rho_i(\boldsymbol{Z}) = \rho_i(Z_i), \quad \text{for all } \boldsymbol{Z} \in \mathcal{L}_1(\mathrm{P},\mathbf{C}).$$

As indicated in (RCP), performance risks are further restricted to the finite-dimensional set $\mathcal{X} \subseteq \mathbf{C}$ and policies are further restricted to the infinite-dimensional set Π. More specific –yet general enough– assumptions on the structure of (RCP) enabling the development of the results advocated in this paper will be discussed in due course.

Problem (RCP) manifests itself in a variety of interesting applications. As a canonical example coming from wireless systems engineering [2], if $i \in \mathbb{N}_N^+$ refers to a user of a wireless system, then the i-th entry of $\boldsymbol{\rho}$ may evaluate the risk associated with the service experienced by the i-th user; hereafter, $\boldsymbol{f}$ will also be called the *service function*. However, with the exception of very special scenarios, the service experienced by the i-th user *can* (and, in general, will) be dependent on the services experienced by the rest of the users in the system. This is due to each service f_i being a function of the coupling variables $\boldsymbol{p}(\boldsymbol{H})$ –the resources– and $\boldsymbol{H}$ –the uncertainty–, which are common to all users in the system. This general structure adheres to, among others, a large variety of resource allocation problems encountered in practice (e.g., see Section 1.1 below, as well as Section 7.1).

Problem (RCP) admits an equivalent and rather useful representation. By the duality theorem for risk measures [1, Theorem 6.5], we have that for every risk measure ρ which is convex, proper, lower semicontinuous and positively homogeneous, it holds that

$$\rho(Z) = \sup_{\zeta \in \mathbb{A}} \langle \zeta, Z \rangle \triangleq \sup_{\zeta \in \mathbb{A}} \int \zeta(\boldsymbol{h}) Z(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}), \quad \text{for all } Z \in \mathcal{L}_1(\mathrm{P},\mathbb{R}),$$

where the *uncertainty set* $\mathbb{A}$ is the domain of the convex conjugate of ρ (i.e., its Legendre-Fenchel transform), and it holds that $\mathbb{A} \subseteq \mathcal{L}_\infty(\mathrm{P},\mathbb{R})$ by functional duality. The set $\mathbb{A}$ is also called the *risk envelope* of ρ. In this paper, we focus on risk envelopes which are subsets of $\{\zeta \in \mathcal{L}_\infty(\mathrm{P},\mathbb{R}) | |\zeta(\cdot)| \leq \gamma, \mathrm{P}\text{-a.e.}\}$, where $\gamma > 0$ is some arbitrarily large but finite constant, but otherwise we make no further assumptions. Note that such a constant always exists whenever ρ is additionally real-valued (and thus continuous) on $\mathcal{L}_1(\mathrm{P},\mathbb{R})$ –i.e., in the class of risk measures appearing in problem (RCP)–, in which case $\mathbb{A}$ coincides with the subdifferential of ρ at the origin, implying that $\mathbb{A}$ is a nonempty and bounded (in fact weakly*-compact) subset of $\mathcal{L}_\infty(\mathrm{P},\mathbb{R})$, and independent of the choice of Z [1, Sections 6.3.1 and 7.3.1]. Of course, it follows that, for every $Z \in \mathcal{L}_1(\mathrm{P},\mathbb{R})$,

$$-\rho(-Z) = \inf_{\zeta \in \mathbb{A}} \langle \zeta, Z \rangle.$$

Under this provisioning, our initial functional program may be equivalently expressed as

$$\boxed{\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq \inf_{\boldsymbol{\zeta} \in \mathbb{A}_\gamma^S} \mathbb{E}\{\boldsymbol{\zeta}(\boldsymbol{H}) \circ \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0} \\ & (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array}}, \qquad \text{(RCP-E)}$$

where "$\circ$" denotes the Hadamard product, the infimum is understood in a component-wise manner, and where the service uncertainty set (i.e., the risk envelope) $\mathbb{A}_\gamma^S$ is defined as the Cartesian product $\mathbb{A}_\gamma^S \triangleq \mathbb{A}_\gamma^1 \times \mathbb{A}_\gamma^2 \times \ldots \times \mathbb{A}_\gamma^N$, with each $\mathbb{A}_\gamma^i$ satisfying the inclusion

$$\mathbb{A}_\gamma^i \subseteq \{\zeta \in \mathcal{L}_\infty(\mathrm{P},\mathbb{R}) \big| |\zeta(\cdot)| \leq \gamma, \mathrm{P}\text{-a.e.}\}, \quad \forall i \in \mathbb{N}_N^+.$$

[1] To avoid confusion throughout, a risk measure is called convex *if and only if* it is a convex functional of its argument; note that this is in contrast to, e.g., [1, Definition 6.4], where a risk measure is called convex if it satisfies additional conditions to mere convexity, namely, monotonicity and translation equivariance.

In passing, it may be worth noting that (RCP-E) is equivalent to the semi-infinite functional program

$$\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \le \mathbb{E}\{\boldsymbol{\zeta}(\boldsymbol{H}) \circ \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\}, \quad \forall \boldsymbol{\zeta} \in \mathbb{A}^S_\gamma\,, \\ & \boldsymbol{g}(\boldsymbol{x}) \ge \mathbf{0} \\ & (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array} \tag{RCP-I}$$

illustrating the generality of the risk-constrained problem (RCP), as compared with its risk-neutral counterpart obtained by choosing $\mathbb{A}^i_\gamma = \{\zeta | \zeta(\cdot) = 1, \text{P-a.e.}\}$ for $i \in \mathbb{N}^+_N$, which is of course equivalent with replacing the risk measure $\boldsymbol{\rho}$ by a finite-dimensional vector of linear functionals (expectations).

1.1 Origins, Related Literature and Contributions

We will be mainly interested in investigating *Lagrangian duality relations* for the infinite-dimensional constrained program (RCP) –in particular, whether strong duality holds or not– under a general set of assumptions, notably without enforcing any additional structure on the integrable service vector function $\boldsymbol{f}(\cdot, \boldsymbol{H})$. Recall that strong duality is a very desirable property for both theoretical and practical purposes and ensures in particular existence of optimal dual variables (Lagrange multipliers), in addition to a null duality gap [3–5]; see Section 3 for a gentle exposition focusing on our purposes herein.

Before we discuss related prior work and existing results, it would be helpful to first present evidence justifying the usefulness and relevance of (RCP) in applications by mainly capitalizing on the standard *risk-neutral* case, i.e., when $\boldsymbol{\rho}$ is a vector of expectations. Such evidence will hopefully provide at least some preliminary motivation for studying (RCP) in the general risk-constrained setting we consider in this paper. Indeed, due to the modular structure of (RCP), and beyond its applicability in numerous settings in resource allocation for networking and wireless communications [2, 6–21], instances of (RCP) appear in areas including nonlinear sparse functional programming [22] with applications such as nonlinear line spectrum estimation and robust functional data analysis [22], nonconvex constrained machine learning [23], the so-called "fluid problem" in dynamic resource allocation [24] (with various applications such as network dynamic pricing, network revenue management, dynamic bidding, online matching, and order fulfillment; see [24] and the references therein), and wireless control [25, 26], to name a few. In all those instances of the constrained problem (RCP), nonconvexity is abundant, and it should be expected that a rigorous dual-domain characterization of (RCP), although highly desirable, is a challenging task, at least under general conditions.

Dual-domain analysis and strong duality results concerning the risk-neutral counterpart of (RCP) just discussed have a relatively short history, starting, to the best of our knowledge, in 2008 with the seminal paper [12] by Luo and Zhang in the context of wireless spectrum management (i.e., resource allocation), which was in turn based on an earlier preliminary discovery reported by Yu and Lui in [27]. Since then, and after subsequent developments by Ribeiro and Giannakis in [2, 14], there has been a flurry of research activity positioned around (RCP) and its various applications, in particular informing an extensive literature on dual-domain methods for obtaining optimal resource policies in the context of networking and wireless systems, including recent advances in model-free learning for wireless communications –see, e.g., references in the previous paragraph–, such as approximate strong duality relations for finite-dimensional versions of (RCP) exploiting convolutional smoothing and universal policy parameterizations [19, 21]. Strong duality of (RCP) in the risk-neutral setting has also been studied in [22], and more recently leveraged in [23] to establish PAC learnability and statistical generalization in the context of constrained machine learning.

More specifically and from a technical standpoint, articles [12] and subsequently [2, 14] were the first to establish the remarkable fact that, under a set of standard and generic assumptions (see our Assumption 1 in Section 4), the most important of them being the *nonatomicity* of the reference Borel measure P, the risk-neutral counterpart of (RCP) exhibits strong duality (also implying a zero duality gap) regardless of the nature of the service $\boldsymbol{f}(\cdot, \boldsymbol{H})$; the latter may in fact be *arbitrary*, in particular (component-wise) nonconcave or even discontinuous. In a sense (see, e.g., [22, Theorem 1 and its proof]), a nonatomic distribution P "patches the holes" of the *image space set* of (RCP) [28], thus preserving its convexity; this follows from a careful application of the celebrated theorem of A. A. Lyapunov [29, Corollary IX.1.6], on the convexity of the range of finite-dimensional nonatomic vector measures. The guaranteed convexity of the image space set of (RCP) can then be utilized to establish strong duality –itself a consequence of convex geometry [30]– of (RCP) in the risk-neutral case via a standard and almost elementary application of the supporting hyperplane theorem. The latter argument is a special part of a longer story about strong duality relations in general conic programming in infinite dimensions, for instance as considered in [30]; see also the books [28, 31].

Evidently, if its risk-neutral version is at all valuable, then so is the substantially more general risk-constrained problem (RCP), which is the focus of this work. This is because most if not all problems discussed above admit at least one interpretable and practically justifiable risk-averse reformulation (by just replacing expectations with appropriate and meaningful risk measures, at the very least). For a concrete example, we refer the reader to our recent article [32], where (RCP) is considered in the "canonical" context of wireless systems resource allocation, albeit under a classical and stylized convex programming setting (thus ensuring strong Lagrangian duality under standard constraint qualifications). In [32], the standard expectation is replaced by the Conditional Value-at-Risk (CVaR) [1, 33], provably resulting in optimal resource policies ensuring robust and reliable system performance, and operationally desirable quality-of-service.

Despite their relevance, though, risk-constrained policy search programs in the form of (RCP) are not currently as well-studied as their risk-neutral versions. We believe that this is natural and mainly due to the fact that, specifically in the nonconvex setting, dual-domain analysis and properties of (RCP), which are crucial for further developments such as the design of efficient algorithms for tackling such problems, are currently absent and essentially unexplored. Our work in this paper is exactly on initiating an effort for rigorously addressing those issues.

Contributions: We show that, perhaps surprisingly, *the risk-constrained functional problem (RCP) exhibits strong duality, under exactly the same assumptions utilized in the risk-neutral case* (i.e., Assumption 1 in Section 4). No further assumptions are required, and the result holds for a wide variety of risk measures, namely all (finite-valued) convex and positively homogeneous risk measures on the space $\mathcal{L}_1$, where the vector $\boldsymbol{\rho}$ may be comprised of different such risk measures in each dimension. Popular risk measures supported within our setting include the CVaR, the mean-absolute deviation (MAD, including the non-monotone case) [34, 35], certain distributionally robust representations and more generally all real-valued coherent risk measures on $\mathcal{L}_1$. Consequently, all members of large classes of coherent risk measures (taken on $\mathcal{L}_1$), such as spectral risk measures [1, Section 6.3.4], or distortion risk measures (with appropriately chosen distortions) [36], or combinations of those, are valid choices for each of the entries of $\boldsymbol{\rho}$ in (RCP).

To the best of our knowledge, the technique we devise to establish strong duality of (RCP) is new. It relies on exploiting risk duality in (RCP) so that its risk constraints, which constitute a finite-dimensional vector of nonlinear functionals, can be *lifted to a vector of linear functionals in infinite dimensions*; see the equivalent problem (RCP-E). The hope is for such a collection of linear functionals to be easier to handle than the risk measures they represent, by leveraging elements of functional analysis, specifically vector measure theory [29].

However, in contrast to the risk-neutral setting, application of Lyapunov's convexity theorem fails in the case of (RCP-E), because the corresponding vector measure construction –in the fashion of [2, 12, 14]– is infinite-dimensional. For such vector measures, Lyapunov's convexity theorem provably may not hold; see, e.g., counter-examples in [29, Section IX. 1]. To bypass this fundamental difficulty, we leverage another well-known result, namely, J. J. Uhl's weak extension [37] of Lyapunov's convexity theorem for vector measures taking values in general infinite-dimensional Banach spaces. This result guarantees convexity *of the norm-closure* of the range of a nonatomic Banach-valued vector measure under certain regularity conditions. Indeed, leveraging Uhl's theorem together with the fact that the functional constraints of problem (RCP-E) are all linear (and under the standard Assumption 1 in Section 4), we are able to prove that *the norm-closure of the image space set* associated with (RCP) (the latter denoted as $\mathcal{C}$) is convex. In a sense, the closure operator further "patches the tears in the fabric" of $\mathcal{C}$, which in the general case might remain as a consequence of the nonlinearity of the risk measures in $\boldsymbol{\rho}$, despite P being nonatomic. We then show that, in fact, this conclusion suffices to establish strong duality of (RCP), as a consequence of convex geometry.

While the strong duality result we develop in this paper is general and expands substantially upon existing literature, it is restricted to real-valued risk measures on $\mathcal{L}_1$. Indeed, extending the result to risk measures whose domains are subsets of $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$, importantly $\mathcal{L}_p(\mathrm{P}, \mathbb{R})$ with $p > 1$, does not seem to be a trivial matter, and we conjecture that such an extension would require a different approach, which we defer to future work. Nonetheless, we may salvage the situation by introducing a CVaR*ization operation* (i.e., an infimal convolution of a given risk measure with the CVaR). We show that any risk measure restricted on $\mathcal{L}_p, p > 1$, can be associated with another –closely related– risk measure that is well-defined and finite on $\mathcal{L}_1$. In fact, we are able to show that the closeness of this new risk measure (finite-valued on $\mathcal{L}_1$) to the original one (finite-valued on $\mathcal{L}_p$) –i.e., the quality of the approximation– can be controlled by the CVaR level of the corresponding CVaR envelope. More specifically, in the coherent case, we show that the family of approximating CVaRized risk measures produced via this technique converges, as the CVaR level goes to zero, to the original (coherent) risk measure in the *Mosco sense* [38], when the domain of the CVaRization is restricted to $\mathcal{L}_p$. As a result, we can obtain closely related approximations of our original problem involving risk measures on some $\mathcal{L}_p$ space, such that strong duality is guaranteed by our theory.

Further extensions of our theory are also discussed, including variations of (RCP) in which risk measures might appear in the objective as well, or cases where the associated linear inequalities might be substituted by appropriate convex conic inequalities. Additionally, our theory provides insights on the efficient frontiers of mean-risk models (see, e.g., [1, Section 6.2] or [39, Section 2]) when the latter are seen as Lagrangian relaxations of constrained programs where the associated dispersion measure (say) is formulated as a constraint; in particular, our main result readily establishes equivalence of the aforementioned programs, in a well-defined sense and under general conditions.

Lastly, we demonstrate the versatility of our base model and assumption system by looking at two particular applications. We first discuss the natural fit of (RCP) in the context of risk-constrained resource allocation in wireless communication systems, and demonstrate the compatibility of (RCP) for two especially relevant models, namely, a multiple access interference channel [19, 21], and a frequency division boradcast channel [2], previously studied in the risk-neutral setting. Subsequently, we consider the setting of functional risk-constrained supervised learning, in which the associated loss functions are allowed to be nonconvex [23]. By imposing standard assumptions compatible with existing literature, we show that strong duality is guaranteed for a wide class of practically relevant nonconvex risk-constrained learning problems. Further, we recover existing strong duality results on risk-neutral constrained learning obtained in [23] but under relaxed assumptions, and enable a unified treatment of risk-constrained regression and classification tasks.

As a remark in passing, we would like to mention that while Lyapunov's convexity theorem has many profound applications in various areas with obvious practical interest, this does not (yet) seem to be the case for its infinite-dimensional extensions, notably by Uhl [37], Knowles [40], and Kadets and Schechtman [41]. Therefore, we believe that the recognition of the usefulness of Uhl's theorem in this paper as a core technical ingredient in proving strong duality of (RCP) in the –hopefully practically relevant– context of risk-constrained policy optimization under a standard Borel space setup is quite interesting, just some fifty-five (55) years after its formulation (Uhl's theorem was published in 1969 [37]).

1.2 Structure and Notation

The structure of the paper is outlined as follows. To better motivate the story and further illuminate the applicability of the setting from a technical perspective, in Section 2 we discuss some of the risk measures covered by our theory resulting in instances of (RCP) of special interest (and its equivalent reformulation (RCP-E)). In Section 3, we briefly introduce standard Lagrangian duality in the context of the functional program (RCP). In Section 4, we state and discuss our assumptions as well as the main result of this work –establishing strong duality of (RCP)–, which we proceed to prove in Section 5. Additional discussion concerning certain extensions, CVaRizations, and implications of our main result are given in Section 6. Then, in Section 7, we briefly elaborate on two practically relevant and general application settings that can be seen as special cases of (RCP), namely risk-constrained wireless systems resource allocation and risk-constrained learning, and finally conclude in Section 8.

Notation –also applicable above–: Bold capital letters (such as $\mathbf{A}$), or calligraphic letters (such as $\mathcal{A}$), or sometimes plain capital letters (such as A) will denote finite-dimensional sets/spaces, such as Euclidean spaces. Double stroke letters (such as $\mathbb{A}$) will denote infinite-dimensional sets/spaces, such as Banach spaces. Math script letters (such as $\mathscr{A}$) will denote σ-algebras. Boldsymbol letters (such as $\boldsymbol{A}$ or $\boldsymbol{a}$) will denote (random) vectors. The space of p-integrable functions from a measurable space $(\Omega, \mathscr{F})$ equipped with a finite measure $\mu : \mathscr{F} \to \mathbb{R}_+$ to a Banach space $\mathbb{A}$, with standard notation $\mathcal{L}_p(\Omega, \mathscr{F}, \mu; \mathbb{A})$ [1, 29], is abbreviated as $\mathcal{L}_p(\mu, \mathbb{A})$, as it is also common practice [29]. The rest of the notation is standard.

2 Risk Measures

As already discussed, problem (RCP) (which is equivalent to (RCP-E) under the conditions mentioned in the introduction) is very general and several cases of special interest can be formulated as particular instances, in addition, of course, to the risk-neutral version of (RCP). We now briefly discuss some standard risk measures and how they fit the adopted framework, as follows.

2.1 Conditional Value-at-Risk

The CVaR at level $\beta \in (0, 1]$ is defined as [1, 33]

$$\text{CVaR}^{\beta}(Z) \triangleq \inf_{t\in\mathbb{R}} t + \frac{1}{\beta}\mathbb{E}\{(Z-t)_+\}, \quad Z \in \mathcal{L}_1(\mathrm{P}, \mathbb{R}),$$

for which (RCP) reduces to

$$\boxed{\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq -\text{CVaR}^{\boldsymbol{\beta}}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \mathbf{0} \\ & (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array}}, \tag{CVaR}$$

where $\boldsymbol{\beta} \triangleq [\beta_1\, \beta_2\, \ldots\, \beta_N] \in (0, 1]^N$ is a vector containing the CVaR levels associated with each entry of the service function $\boldsymbol{f}$. In this case, the equivalent formulation of (CVaR) in the form of (RCP-E) is valid by choosing $\gamma = \max_{i\in\mathbb{N}_N^+} 1/\beta_i$, and with corresponding risk envelopes given by [1, Example 6.19]

$$\mathbb{A}_{\gamma}^{i} = \{\zeta \in \mathcal{L}_\infty(\mathrm{P}, \mathbb{R}) | \zeta(\cdot) \in [0, 1/\beta_i], \text{P-a.e., and } \mathbb{E}\{\zeta\} = 1\}, \quad \forall i \in \mathbb{N}_N^+.$$

Note that in the case of CVaR, the resulting functional program may also be stated as

$$\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq \sup_{\boldsymbol{t}\in\mathbf{C}} \boldsymbol{t} + \dfrac{1}{\boldsymbol{\beta}} \circ \mathbb{E}\{-(\boldsymbol{t} - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}))_+\} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \mathbf{0} \\ & (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array},$$

where each risk measure of the corresponding risk constraint amounts to the "reward version" of the CVaR, the latter being usually defined for minimizing costs. This problem is of course equivalent to

$$\begin{array}{ll} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot),\boldsymbol{t}}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq \boldsymbol{t} + \dfrac{1}{\boldsymbol{\beta}} \circ \mathbb{E}\{-(\boldsymbol{t} - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}))_+\} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \mathbf{0} \\ & (\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{t}) \in \mathcal{X} \times \Pi \times \mathbf{C} \end{array}.$$

As discussed in Section 1.1, this problem has been recently considered in [32] to discover optimal risk-aware resource allocation policies in the context of wireless systems, albeit in a convex programming framework, where $\boldsymbol{f}(\cdot, \boldsymbol{H})$ is of special form and in particular component-wise concave. Our work in this paper essentially extends strong duality of (CVaR) in the case where $\boldsymbol{f}(\cdot, \boldsymbol{H})$ is abritrary (and merely integrable).

2.2 Mean-Absolute Deviation

Another popular special case is that of the Mean-Absolute Deviation with trade-off parameter $\lambda \geq 0$ (i.e., monotone or not) defined as [34, 35]

$$\mathrm{MAD}^{\lambda}(Z) \triangleq \mathbb{E}\{Z\} + \lambda \mathbb{E}\{|Z - \mathbb{E}\{Z\}|\}, \quad Z \in \mathcal{L}_1(\mathrm{P}, \mathbb{R}),$$

for which problem (RCP) reduces to

$$\boxed{\begin{array}{ll} \underset{\boldsymbol{x}, \boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq -\mathrm{MAD}^{\boldsymbol{\lambda}}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0} \\ & (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array}}, \tag{MAD}$$

where $\boldsymbol{\lambda} \triangleq [\lambda_1\, \lambda_2\, \ldots\, \lambda_N] \in \mathbb{R}_+^N$ is another vector containing the MAD trade-offs associated with each entry of the service function $\boldsymbol{f}$. In this case, (MAD) can be equivalently written in the form of (RCP-E) by choosing $\gamma = \max_{i \in \mathbb{N}_N^+} 1 + 2\lambda_i$, and risk envelopes [1, Example 6.22]

$$\mathbb{A}_\gamma^i = \{\zeta \in \mathcal{L}_\infty(\mathrm{P}, \mathbb{R}) | \zeta = 1 + \zeta' - \mathbb{E}\{\zeta'\}, \text{ and } \|\zeta'\|_{\mathcal{L}_\infty} \leq \lambda_i\}, \quad \forall i \in \mathbb{N}_N^+.$$

Similar to the case of CVaR, it is easy to see that the MAD program takes the form

$$\begin{array}{ll} \underset{\boldsymbol{x}, \boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq \mathbb{E}\{\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} - \boldsymbol{\lambda} \circ \mathbb{E}\{|\mathbb{E}\{\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})|\} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0} \\ & (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array},$$

or, equivalently,

$$\begin{array}{ll} \underset{\boldsymbol{x}, \boldsymbol{p}(\cdot), \boldsymbol{t}}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq \boldsymbol{t} - \boldsymbol{\lambda} \circ \mathbb{E}\{|\boldsymbol{t} - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})|\} \\ & \boldsymbol{t} = \mathbb{E}\{\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0} \\ & (\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{t}) \in \mathcal{X} \times \Pi \times \mathbf{C} \end{array}.$$

This paper shows that the (MAD) problem exhibits strong duality for any choice of the weight vector $\boldsymbol{\lambda} \geq 0$, and for a merely integrable and otherwise arbitrary $\boldsymbol{f}(\cdot, \boldsymbol{H})$.

2.3 Coherent Risk Measures on $\mathcal{L}_1$

Generalizing, there are numerous other risk measures which are compatible with our assumptions on the vector risk measure $\boldsymbol{\rho}$. In fact, all coherent risk measures on the space $\mathcal{L}_1$ (relative to any qualifying choice of the probability measure P) are supported under the adopted framework. For any such risk measure ρ, it is well known that

$$\rho(Z) = \sup_{\zeta \in \mathbb{A}} \langle \zeta, Z \rangle = \sup_{\frac{\mathrm{dQ}}{\mathrm{dP}} \in \mathbb{A}} \mathbb{E}\left\{ Z(\boldsymbol{H}) \frac{\mathrm{dQ}}{\mathrm{dP}}(\boldsymbol{H}) \right\}, \quad \text{for all } Z \in \mathcal{L}_1(\mathrm{P}, \mathbb{R}),$$

where the risk envelope $\mathbb{A}$ takes the special form [1, Section 6.3]

$$\begin{aligned}\mathbb{A} &= \{\zeta \in \mathcal{L}_\infty(\mathrm{P}, \mathbb{R}) | \langle \zeta, Z \rangle \leq \rho(Z) \text{ for all } Z \in \mathcal{L}_1(\mathrm{P}, \mathbb{R}), \zeta \succeq 0, \mathbb{E}\{\zeta\} = 1\} \\ &= \left\{ \frac{\mathrm{dQ}}{\mathrm{dP}} \in \mathcal{L}_\infty(\mathrm{P}, \mathbb{R}) \middle| \left\langle \frac{\mathrm{dQ}}{\mathrm{dP}}, Z \right\rangle \leq \rho(Z), \text{ for all } Z \in \mathcal{L}_1(\mathrm{P}, \mathbb{R}) \right\}.\end{aligned}$$

In the above, $\mathrm{dQ}/\mathrm{dP} = \zeta$ denotes the Radon-Nikodym derivative of a probability measure Q on $(\mathcal{H}, \mathscr{B}(\mathcal{H}))$ relative to P, with the former assumed to be absolutely continuous relative to the latter; we use the standard notation $\mathrm{Q} \ll \mathrm{P}$.

If ρ is also real-valued, it is continuous in the strong topology on $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$, and its risk envelope $\mathbb{A}$ is consistent with our setting, being in particular a convex bounded and weakly*-closed (in fact weakly*-compact) subset of the set of densities in $\mathcal{L}_\infty(\mathrm{P}, \mathbb{R})$ (and independent of each choice of Z). It follows that ρ admits a distributionally robust representation of the form

$$\rho(Z) = \sup_{\mathrm{Q} \in \mathfrak{M}} \mathbb{E}_\mathrm{Q}\{Z(\boldsymbol{H})\}, \quad \text{for all } Z \in \mathcal{L}_1(\mathrm{P}, \mathbb{R}),$$

where

$$\begin{aligned}\mathfrak{M} &\triangleq \left\{ \mathrm{Q} \ll \mathrm{P} \middle| \left\langle \frac{\mathrm{dQ}}{\mathrm{dP}}, Z \right\rangle \leq \rho(Z), \text{ for all } Z \in \mathcal{L}_1(\mathrm{P}, \mathbb{R}), \text{ with } \frac{\mathrm{dQ}}{\mathrm{dP}}(\cdot) \leq \gamma, \mathrm{P}\text{-a.e.} \right\} \\ &\subseteq \left\{ \mathrm{Q} \ll \mathrm{P} \middle| \mathrm{Q} \leq \gamma \mathrm{P}, \text{ on Borel sets} \right\}.\end{aligned}$$

To further highlight the versatility of (RCP), we would like to emphasize that one can freely "mix-and-match" any combination of risk measures present in the constraints of (RCP), out of the large variety of those supported under our assumptions. Such combinations include different choices of risk measures across the components of the service vector $\boldsymbol{f}$, e.g., expectations for some components and CVaRs or MADs for others, as well as combinations of different risk measures for each of the components of $\boldsymbol{f}$; a classical example is that of mean-CVaR trade-offs.

3 Lagrangian Duality

A celebrated approach for dealing with the explicit inequality constraints of the risk-aware problem (RCP) is by exploiting *Lagrangian duality*, which has been proven fundamental in analyzing and efficiently solving constrained convex optimization problems; see, e.g., [3–5]. Note, however, that since the services $\boldsymbol{f}(\cdot, \boldsymbol{H})$ appearing in problem (RCP) are nonconcave in general (i.e., with respect to the first argument corresponding to the policy $\boldsymbol{p}$), standard results in Lagrangian duality for convex optimization do not automatically apply. On top of that, one has to incorporate the structural complexity of the risk measure $\boldsymbol{\rho}$, which is a nonlinear functional of its argument. As a result, fundamental properties of expectation (being the most trivial risk measure) which enable an elegant and straightforward analysis, such as linearity, do not hold for $\boldsymbol{\rho}$.

The *Lagrangian function* $\mathsf{L} : \mathbf{C} \times \Pi \times \mathbb{R}^{N_g} \times \mathbf{C} \to \mathbb{R}$ associated with the risk-constrained problem (RCP) is defined by scalarizing its constraints as

$$\mathsf{L}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}) \triangleq g^o(\boldsymbol{x}) + \langle \boldsymbol{\lambda_g}, \boldsymbol{g}(\boldsymbol{x}) \rangle + \left\langle \boldsymbol{\lambda_\rho}, -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) - \boldsymbol{x} \right\rangle,$$

where $\boldsymbol{\lambda} \equiv (\boldsymbol{\lambda_g}, \boldsymbol{\lambda_\rho}) \in \mathbb{R}^{N_g} \times \mathbf{C}$ are dual multipliers associated with the constraints of (RCP). Then the *dual function* $\mathsf{D} : \mathbb{R}^{N_g} \times \mathbf{C} \to (-\infty, \infty]$ is defined as

$$\mathsf{D}(\boldsymbol{\lambda}) \triangleq \sup_{(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi} \mathsf{L}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}).$$

If the optimal value of problem (RCP) is $\mathsf{P}^* \in \mathbb{R}$, it is then easily understood that $\mathsf{P}^* \leq \mathsf{D}$ on the positive orthant (i.e., for $\boldsymbol{\lambda} \geq \boldsymbol{0}$), and thus it is most reasonable to consider the *dual problem*

$\inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \mathsf{D}(\boldsymbol{\lambda})$, which is always convex and whose optimal value

$$\mathsf{D}^* \triangleq \inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \mathsf{D}(\boldsymbol{\lambda}) \in (-\infty, \infty]$$

serves as the tightest over-estimate of the optimal value of problem (RCP), P^*, when knowing only D. Then, one of the basic questions in Lagrangian duality is whether we can essentially replace an original constrained problem with its dual, in the sense that $\mathsf{P}^* \equiv \mathsf{D}^*$; in such a case, we say that the problem has *zero duality gap*. Referring to problem (RCP), this would imply that it can be replaced by the *minimax problem*

$$\inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \mathsf{D}(\boldsymbol{\lambda}) \equiv \inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \sup_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi} \mathsf{L}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\lambda}),$$

whose optimal value is D^*. If D^* is attained, i.e., an optimal multiplier vector exists, then (RCP) is said to exhibit *strong duality*. A zero duality gap also implies that problem (RCP) satisfies the *saddle point property* (whether a saddle point exists or not), which is expressed as

$$\mathsf{D}^* = \inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \sup_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi} \mathsf{L}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\lambda}) = \sup_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi} \inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \mathsf{L}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\lambda}) = \mathsf{P}^*.$$

Zero duality gaps are desirable: Because there is a finite number of constraints, the dual function is finite-dimensional even though the original functional problem (RCP) is infinite-dimensional. Additionally, for every choice of the dual variable $\boldsymbol{\lambda}\geq \mathbf{0}$ (and therefore for any optimal multiplier vector), joint maximization of the Lagrangian $\mathsf{L}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\lambda})$ over the pair $(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi$ is separable. We thus see that duality transforms a constrained problem into an unconstrained problem in a principled and favorable way and, provided that the original constrained stochastic program (in our case (RCP)) exhibits zero duality gap or strong duality, presents a general methodological approach to tackle it. While zero duality gaps are a common and fundamental characteristic of problems in convex (concave) constrained optimization (under appropriate regularity conditions), proving zero duality gaps in nonconvex problems such as (RCP) is a much more delicate and challenging task.

4 Main Result

Quite surprisingly, the set of assumptions enabling the dual-domain analysis of problem (RCP) that we will develop is standard and exactly the same as compared with the relevant literature on the corresponding risk-neutral related problems; see, e.g., [2, 12, 22, 23] and also Section 1.1.

Assumption 1 (Problem Setting). *The following conditions are in effect:*

1. *The utilities g^o and $\boldsymbol{g}$ are concave.*
2. *The feasible set $\mathcal{X}$ of ergodic services is convex.*
3. *The policy feasible set Π is decomposable.*
4. *The Borel measure P is nonatomic[a].*
5. *Problem (RCP) satisfies Slater's condition (i.e., inequality constraints are strictly feasible).*

[a] Recall that P is nonatomic if for any event E with $\mathrm{P}(E) > 0$, an event $E' \subseteq E$ exists such that $\mathrm{P}(E) > \mathrm{P}(E') > 0$.

A useful observation is that the conditions comprising Assumption 1 are fully compatible, and can be easily satisfied in practice. For instance, the space of all (Borel-)measurable functions (policies) from $\mathcal{H}$ to $\mathbf{R}$ as well as all spaces of integrable functions $\mathcal{L}_p(\mathrm{P},\mathbf{R})$, $p\in[1,\infty]$, are decomposable [42], and all those examples can be possible choices for the feasible set Π. A more specific standard choice of a decomposable feasible set Π, which is relevant in applications, is the uniform box

$$\Pi = \left\{\boldsymbol{p} : \mathcal{H} \to \mathbf{R} \middle| \operatorname{ess\,sup}_{\mathrm{P}} \|\boldsymbol{p}(\cdot)\|_\infty \leq U\right\},$$

where $U > 0$ is an appropriate fixed number, or, the more refined rectangular box

$$\Pi = \left\{\boldsymbol{p} : \mathcal{H} \to \mathbf{R} \middle| \operatorname{ess\,sup}_{\mathrm{P}} p^i(\cdot) \leq U^i, i\in\mathbb{N}^+_{N_{\boldsymbol{p}}}\right\},$$

where the U^i's are fixed. In a more delicate setting, another also standard choice for a decomposable Π is (see, e.g., [2] and references therein)

$$\Pi = \{\boldsymbol{p} : \mathcal{H} \to \mathbf{R} | \boldsymbol{p}(\boldsymbol{h}) \in \mathcal{U}(\boldsymbol{h}), \text{P-a.e.}\},$$

where $\mathcal{U} : \mathcal{H} \rightrightarrows \mathbf{R}$ is a closed-valued multifunction, which is *also* closed [1, Section 7.2.3]; this implies in particular that $\mathcal{U}$ admits at least one measurable selection [1, Theorem 7.39]. In other words, every feasible $\boldsymbol{p} \in \Pi$ may be taken to be a well-behaved Borel-measurable selection of $\mathcal{U}$. Regarding the rest of the conditions of Assumption 1, all are standard as mentioned above. We would just like to further point out that condition (4) holds naturally if the Borel measure P has a density with respect with the Lebesgue measure; this is a valid assumption in numerous practical settings.

We are now in position to state the main result of this work. The detailed proof is presented next in Section 5.

Theorem 1 (**Strong Duality in Risk-Constrained Nonconvex Functional Programming**). *Let Assumption 1 be in effect. Then problem (RCP) has zero duality gap, i.e.,* $\mathsf{P}^* = \mathsf{D}^*$*. In fact, (RCP) exhibits strong duality, i.e., optimal dual variables exist.*

5 Proof of Theorem 1

Hereafter, we let Assumption 1 be in effect. We will work with the utility-constraint set –also called an image space set [28, 30]– associated with problem (RCP) and defined as

$$\mathcal{C} \triangleq \left\{ (\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \,\middle|\, \begin{array}{l} g^o(\boldsymbol{x}) \geq \delta_o \\ -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) - \boldsymbol{x} \geq \boldsymbol{\delta}_r \,, \text{ for some } (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \\ \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{\delta}_d \end{array} \right\}.$$

Following [22, 23], showing that $\mathcal{C}$ is convex and the strict feasibility of problem (RCP) (i.e., Slater's condition in Assumption 1) would suffice to ensure strong duality of (RCP), as a relatively simple consequence of the supporting hyperplane theorem; see Section 5.3, or [22, Theorem 1 and its proof], or [23, Appendix A] for the details. In fact, this is a special part of a far more general story; see, e.g., [30]. Proving convexity of $\mathcal{C}$ is nontrivial in the case of (RCP) though, and does not follow from the analyses presented in the aforementioned articles. This is due to the nonlinearity of the functionals present in the risk constraints of (RCP), in sharp contrast to standard problems considered in the literature (e.g., in [2, 12, 22, 23]), where the corresponding constraints evaluate the vector $\boldsymbol{f}$ solely through linear functionals, i.e., expectations.

The Challenge of Risk: From the discussion above it follows that ultimately we would like to prove that the set $\mathcal{C}$ is convex. This would mean that if $(\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \in \mathcal{C}$ for $(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi$ and if $(\delta'_o, \boldsymbol{\delta}'_r, \boldsymbol{\delta}'_d) \in \mathcal{C}$ for $(\boldsymbol{x}', \boldsymbol{p}') \in \mathcal{X} \times \Pi$, then, for every $\alpha \in [0, 1]$, it should be the case that

$$\alpha(\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) + (1 - \alpha)(\delta'_o, \boldsymbol{\delta}'_r, \boldsymbol{\delta}'_d) \in \mathcal{C}.$$

In other words, we would have to show that there exists another pair $(\boldsymbol{x}_\alpha, \boldsymbol{p}_\alpha) \in \mathcal{X} \times \Pi$, such that

$$\begin{aligned} g^o(\boldsymbol{x}_\alpha) &\geq \alpha\delta_o + (1 - \alpha)\delta'_o, \\ -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha(\boldsymbol{H}), \boldsymbol{H})) - \boldsymbol{x}_\alpha &\geq \alpha\boldsymbol{\delta}_r + (1 - \alpha)\boldsymbol{\delta}'_r \quad \text{and} \\ \boldsymbol{g}(\boldsymbol{x}_\alpha) &\geq \alpha\boldsymbol{\delta}_d + (1 - \alpha)\boldsymbol{\delta}'_d. \end{aligned}$$

By choosing $\boldsymbol{x}_\alpha = \alpha\boldsymbol{x} + (1 - \alpha)\boldsymbol{x}' \in \mathcal{X}$ (by assumption, $\mathcal{X}$ is convex), and appealing to the concavity of g^o and $\boldsymbol{g}$, the proof would be complete if we showed that, for every $\alpha \in [0, 1]$, there is a policy $\boldsymbol{p}_\alpha \in \Pi$, such that

$$-\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha(\boldsymbol{H}), \boldsymbol{H})) \geq -\alpha\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) - (1 - \alpha)\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})). \quad (1)$$

Unfortunately, proving that $\mathcal{C}$ is convex for the risk-constrained setting seems impossible, and we conjecture that such an assertion is probably false, at least unless we enforce additional structural conditions (on top of the fully compatible set of conditions comprising Assumption 1).

Fortunately though, what we can indeed show is *convexity of the closure of* $\mathcal{C}$; establishing this key result constitutes the core of the proof of Theorem 1. Then, convexity of $\text{cl}(\mathcal{C})$ is combined with the fact that the point $(\mathsf{P}^*, \mathbf{0}, \mathbf{0})$ cannot be in the interior of $\text{cl}(\mathcal{C})$, enabling a separating hyperplane argument for the pair $(\mathsf{P}^*, \mathbf{0}, \mathbf{0})$ and $\text{cl}(\mathcal{C})$, which together with Slater's condition (i.e., strict feasibility) implies strong duality for problem (RCP); this step is similar to that in the risk-neutral case, where $\mathcal{C}$ is a convex set, see, e.g., [22]. We believe that establishing strong duality for (RCP) using merely convexity of $\text{cl}(\mathcal{C})$ is interesting, as it reveals a substantially weaker chain of self-contained logical arguments sufficient to prove Theorem 1 as compared with the risk-neutral setting (a special case). For an alternative, though non-elementary and more involved way of establishing strong duality for (RCP) –again by exploiting convexity of $\text{cl}(\mathcal{C})$ and Slater's condition–, see Appendix C.

In passing, note that (1) is trivially true and thus convexity of $\mathcal{C}$ is implied in the special case where $\boldsymbol{f}(\cdot, \boldsymbol{h})$ is concave for every $\boldsymbol{h} \in \mathcal{H}$ and Π *is a convex set* (instead of a decomposable one), simply by choosing $\boldsymbol{p}_\alpha$ as a convex combination of $\boldsymbol{p}$ and $\boldsymbol{p}'$, and exploiting convexity and positive homogeneity of the vector risk measure $\boldsymbol{\rho}$; this is a standard case of convex programming.

5.1 Preliminaries on Vector Measures in Banach Spaces

Let us first introduce some basic definitions and notation from the study of vector measures taking values in general, infinite-dimensional Banach spaces. For a comprehensive treatment of the subject, the reader is referred to the classical monograph [29].

Suppose that $\mathbb{X}$ is a possibly infinite-dimensional Banach space. A *vector measure* on a measurable space $(\Omega, \mathscr{F})$ is a function $\boldsymbol{G} : \mathscr{F} \to \mathbb{X}$. A vector measure $\boldsymbol{G}$ is called countably additive (in the norm topology of $\mathbb{X}$) in the same fashion as a regular real-valued measure. The *variation* of a vector measure $\boldsymbol{G}$ is another function on sets $|\boldsymbol{G}| : \mathscr{F} \to \mathbb{R}_+$ defined as

$$|\boldsymbol{G}|(E) \triangleq \sup_{\pi \text{ is a finite partition of } E} \sum_{A \in \pi} \|\boldsymbol{G}(A)\|_{\mathbb{X}}.$$

Then $\boldsymbol{G}$ is suggestively said to be of *bounded variation* whenever $|\boldsymbol{G}|(\Omega) < \infty$. A vector measure $\boldsymbol{G}$ is called *nonatomic* if every event $E \in \mathscr{F}$ such that $\boldsymbol{G}(E) \neq \mathbf{0}$ can be partitioned into events E' and $E \backslash E'$ such that $\boldsymbol{G}(E') \neq \mathbf{0}$ and $\boldsymbol{G}(E \backslash E') \neq \mathbf{0}$; in other words, every event of non-zero measure can be split into two events of non-zero measure.

Further, in the following we will be using the concept of a *Bochner integral*, which is a now standard extension of the Lebesgue integral for functions taking values in infinite-dimensional Banach spaces. While we do not provide a formal description here, the reader is referred to the excellent exposition in [29, Section II], which is a standard textbook on the subject.

Lastly, our analysis will be based on the following extension to the celebrated convexity theorem of A. A. Lyapunov, due to Uhl [37, Theorem 1 and last paragraph before the References section], see also [29, Theorem IX.1.10]. This also classical result conveniently generalizes the convexity theorem to infinite-dimensional Banach spaces, albeit with some nontrivial provisioning on the topological properties of the range of the involved vector measure.

Theorem 2 (**[37] Weak Lyapunov Theorem for the Strong Topology**). *Let* $(\Omega, \mathscr{F})$ *be a measurable space, and let* $\mathbb{X}$ *be any Banach space. Let* $\boldsymbol{G} : \mathscr{F} \to \mathbb{X}$ *be a countably additive vector measure of bounded variation. If* $\boldsymbol{G}$ *is nonatomic and admits a Radon-Nikodym representation, i.e., there exist a finite measure* $\mu : \mathscr{F} \to \mathbb{R}_+$ *and a function* $\boldsymbol{f} \in \mathcal{L}_1(\mu, \mathbb{X})$ *such that*

$$\boldsymbol{G}(E) = \int_E \boldsymbol{f}(\omega) \mathrm{d}\mu(\omega), \quad E \in \mathscr{F},$$

then the norm closure of the range $\boldsymbol{G}(\mathscr{F})$ *is convex and norm-compact.*

5.2 Core of the Proof: Convexity of cl($\mathcal{C}$)

We identify with $\mathbb{X}$ the Banach space of all real-valued sequences bounded in the sup norm, i.e., $\mathbb{X} = \ell_\infty$. For a feasible policy $\boldsymbol{p} \in \Pi$, provisionally define the vector measure $\boldsymbol{G_p} : \mathscr{B}(\mathcal{H}) \to \mathbb{X}$ as

$$\boldsymbol{G_p}(E) \triangleq \begin{bmatrix} \int_E \lambda_0(\boldsymbol{h})\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{h})\mathrm{dP}(\boldsymbol{h}) \\ \int_E \lambda_1(\boldsymbol{h})\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{h})\mathrm{dP}(\boldsymbol{h}) \\ \vdots \\ \int_E \lambda_n(\boldsymbol{h})\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{h})\mathrm{dP}(\boldsymbol{h}) \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})\lambda_0(\boldsymbol{H})\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\} \\ \mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})\lambda_1(\boldsymbol{H})\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\} \\ \vdots \\ \mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})\lambda_n(\boldsymbol{H})\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\} \\ \vdots \end{bmatrix}, \; E \in \mathscr{B}(\mathcal{H}),$$

where $\mathbb{B} \triangleq \{\lambda_n\}_{n\in\mathbb{N}}$ is a countable base (i.e., a dense subset) on $\mathcal{L}_1(\mathrm{P},\mathbb{R})$, for convenience consisting of simple functions on disjoint dyadic cubes on $\mathcal{H}$ with rational coefficients (the construction of such a dense subset is a standard procedure; see, e.g., [43, Chapter 13]). We will be later using the base $\mathbb{B}$ to approximate elements in $\mathbb{A}_\gamma^i$, $i \in \mathbb{N}_N^+$, each of which is a bounded subset of $\mathcal{L}_\infty(\mathrm{P},\mathbb{R})$, and thus of $\mathcal{L}_1(\mathrm{P},\mathbb{R})$ (P is finite); therefore, without loss of generality we may very well assume that $|\lambda_n| \le \gamma$ everywhere on $\mathcal{H}$, and we do so hereafter (otherwise, just take any qualifying countable base on $\mathcal{L}_1(\mathrm{P},\mathbb{R})$ and project each of its members onto the box $[-\gamma,\gamma]$ –choose $\gamma \in \mathbb{Q}$ if needed and note that a clipped simple function is itself a simple function–; then, for every $\zeta \in \mathbb{A}_\gamma^i$, there exists a subsequence $\{\lambda_n\}_{n\in\mathcal{K}}, \mathcal{K} \subseteq \mathbb{N}$ converging to ζ in $\mathcal{L}_1$, and in fact

$$0 \le \|\mathrm{proj}_{[-\gamma,\gamma]}(\lambda_n) - \zeta\|_{\mathcal{L}_1} = \|\mathrm{proj}_{[-\gamma,\gamma]}(\lambda_n) - \mathrm{proj}_{[-\gamma,\gamma]}(\zeta)\|_{\mathcal{L}_1} \le \|\lambda_n - \zeta\|_{\mathcal{L}_1} \underset{n\to\infty}{\longrightarrow} 0,$$

due to nonexpansiveness of the projection map). It is then guaranteed that, for every $E \in \mathscr{B}(\mathcal{H})$, all entries of the vector $\boldsymbol{G_p}(E)$ are well-defined and finite.

For every $E \in \mathscr{B}(\mathcal{H})$, it also readily follows that (below the ordinary absolute value $|\cdot|$ is taken component-wise when presented with a vector as its input)

$$\begin{aligned} \|\boldsymbol{G_p}(E)\|_{\ell_\infty} &= \sup_{n\in\mathbb{N}} \|\mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})\lambda_n(\boldsymbol{H})\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\}\|_\infty \\ &\le \sup_{n\in\mathbb{N}} \|\mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})|\lambda_n(\boldsymbol{H})||\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})|\}\|_\infty \\ &\le \gamma\|\mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})|\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})|\}\|_\infty < \infty, \end{aligned}$$

verifying that $\boldsymbol{G_p}(E)$ is an element of $\mathbb{X}$ for every qualifying E. Further, we can show that, by its construction, $\boldsymbol{G_p}$ can be represented as a Bochner integral as (this is nontrivial; see Appendix A for a detailed verification)

$$\boldsymbol{G_p}(E) = \int_E \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}),\boldsymbol{h})\mathrm{dP}(\boldsymbol{h}), \quad E \in \mathscr{B}(\mathcal{H}),$$

where "$\otimes$" denotes the Kronecker product, and where one can verify that $\boldsymbol{\Lambda}_{\mathbb{B}} \triangleq [\lambda_0\,\lambda_1\,\ldots\,\lambda_n\,\ldots] \in \mathbb{X}$ (see, e.g., [29, Example II.2.10]); note that $\boldsymbol{\Lambda}_{\mathbb{B}}(\cdot) \otimes \boldsymbol{f}(\boldsymbol{p}(\cdot),\cdot)$ is $\mathbb{X}$-valued and Bochner integrable as well (follows from [29, Theorem II.2.2] –again, see Appendix A for details–, also see proof of Lemma 3 below). Then, it follows that $\boldsymbol{G_p}$ is both countably additive and of bounded variation [29, Theorem II.2.4 (iii) and (iv)]. We also use the familiar probabilistic notation

$$\boldsymbol{G_p}(E) = \mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{H}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\}, \quad E \in \mathscr{B}(\mathcal{H}),$$

with an understanding that expectation here is in the Bochner sense (i.e., expectation of a random element taking values in an infinite-dimensional Banach space).

Using the construction above, and together with another feasible policy $\boldsymbol{p}' \in \Pi$, we define another vector measure $\boldsymbol{G} : \mathscr{B}(\mathcal{H}) \to \mathbb{X}$ as

$$\boldsymbol{G}(E) \triangleq \mathbb{E}\left\{\mathbb{1}_E(\boldsymbol{H})\,\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{H}) \otimes \begin{bmatrix} \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}) \\ \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}),\boldsymbol{H}) \end{bmatrix}\right\}, \quad E \in \mathscr{B}(\mathcal{H}),$$

which will serve as our main construction for the rest of the analysis. The next key result is concerned with the structure of the range of $\boldsymbol{G}$, with its proof also verifying that $\boldsymbol{G}$ is essentially an interleaved concatenation of the vector measures $\boldsymbol{G}_{\boldsymbol{p}}$ and $\boldsymbol{G}_{\boldsymbol{p}'}$, as initially defined above.

Lemma 3. *The norm (strong) closure of the range of* $\boldsymbol{G}$

$$\overline{\boldsymbol{G}(\mathscr{B}(\mathcal{H}))} = \overline{\{\boldsymbol{x} \in \mathbb{X} | \boldsymbol{x} = \boldsymbol{G}(E),\ \text{for some } E \in \mathscr{B}(\mathcal{H})\}}$$

is convex and norm-compact.

Proof of Lemma 3. We need to verify the conditions under which Theorem 2 (Uhl) is valid. First, it is true that

$$\begin{aligned}
\mathbb{E}\left\{\left\|\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{H}) \otimes \begin{bmatrix} \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}) \\ \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}),\boldsymbol{H}) \end{bmatrix}\right\|_{\ell_\infty}\right\} &= \mathbb{E}\left\{\sup_{n\in\mathbb{N}}\left\|\lambda_n(\boldsymbol{H}) \begin{bmatrix} \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}) \\ \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}),\boldsymbol{H}) \end{bmatrix}\right\|_{\infty}\right\} \\
&\le \mathbb{E}\left\{\sup_{n\in\mathbb{N}}|\lambda_n(\boldsymbol{H})|\left\|\begin{bmatrix} \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}) \\ \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}),\boldsymbol{H}) \end{bmatrix}\right\|_{\infty}\right\} \\
&\le \gamma\mathbb{E}\left\{\left\|\begin{bmatrix} \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}) \\ \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}),\boldsymbol{H}) \end{bmatrix}\right\|_{\infty}\right\} \\
&\le \gamma\mathbb{E}\left\{\left\|\begin{bmatrix} \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}) \\ \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}),\boldsymbol{H}) \end{bmatrix}\right\|_{1}\right\} < \infty.
\end{aligned}$$

This shows that

$$\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{H}) \otimes \begin{bmatrix} \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}) \\ \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}),\boldsymbol{H}) \end{bmatrix} \in \mathcal{L}_1(\mathrm{P},\mathbb{X}),$$

which implies that $\boldsymbol{G}$ is countably additive [29, Theorem II.2.4 (iii)], has bounded variation [29, Theorem II.2.4 (iv)] and evidently is Radon-Nikodym representable.

To show that $\boldsymbol{G}$ is nonatomic, consider its primitive construction (this is justified in the same way as for $\boldsymbol{G}_{\boldsymbol{p}}$; also see Appendix A) and suppose that $E \in \mathscr{B}(\mathcal{H})$ is such that

$$\boldsymbol{G}(E) \neq \boldsymbol{0} \iff \mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})\lambda_n(\boldsymbol{H})f_i(\widetilde{\boldsymbol{p}}(\boldsymbol{H}),\boldsymbol{H})\} \neq 0, \text{ for some } n \in \mathbb{N}, i \in \mathbb{N}_N^+ \text{ and for } \widetilde{\boldsymbol{p}} = \boldsymbol{p} \text{ or } \boldsymbol{p}'.$$

Note that we necessarily have $\mathrm{P}(E) > 0$ (for otherwise $\boldsymbol{G}(E) = \boldsymbol{0}$). Without loss of generality take $\widetilde{\boldsymbol{p}} = \boldsymbol{p}$ and $n = i = 1$. Then, by a lemma of Blackwell [44, Lemma], nonatomicity of P implies the existence of a Borel subset $\widetilde{E} \subseteq \mathcal{H}$ such that $\mathrm{P}(\widetilde{E}) = \mathrm{P}(\mathcal{H})/2 = 1/2$, for which

$$\mathbb{E}\{\mathbb{1}_{E\cap\widetilde{E}}(\boldsymbol{H})\lambda_1(\boldsymbol{H})f_1(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\} = \frac{1}{2}\mathbb{E}\{\mathbb{1}_E(\boldsymbol{H})\lambda_1(\boldsymbol{H})f_1(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})\} \neq 0.$$

Letting $E' \triangleq E\bigcap\widetilde{E}$, this necessarily implies that $\boldsymbol{G}(E') \neq \boldsymbol{0}$ as well as $\boldsymbol{G}(E\backslash E') \neq \boldsymbol{0}$; in fact, observe that necessarily $E' \subset E$ and that $\mathrm{P}(E') > 0$. By definition, it follows that $\boldsymbol{G}$ has no atoms. Consequently, the conditions of Theorem 2 are fulfilled. Enough said. ■

The conclusions of Lemma 3, in particular convexity of norm closure of the range of $\boldsymbol{G}$, are sufficient to ensure convexity of the (norm) closure of $\mathcal{C}$. To see this, let us first consider the range $\boldsymbol{G}(\mathscr{B}(\mathcal{H}))$ of $\boldsymbol{G}$. Of course $\boldsymbol{z} = \boldsymbol{G}(\mathcal{H})$ and $\boldsymbol{z}' = \boldsymbol{G}(\emptyset) = \boldsymbol{0}$ are both elements of $\boldsymbol{G}(\mathscr{B}(\mathcal{H}))$. Therefore, Lemma 3 implies that, for every $\alpha \in [0,1]$, the convex combination $\alpha\boldsymbol{z} + (1-\alpha)\boldsymbol{z}' \equiv \alpha\boldsymbol{z}$ lies in the norm closure of $\boldsymbol{G}(\mathscr{B}(\mathcal{H}))$. In other words, for each $\alpha \in [0,1]$, there exists a sequence of events $\{E_n^\alpha \in \mathscr{B}(\mathcal{H})\}_{n\in\mathbb{N}}$ such that

$$\lim_{n\to\infty} \|\alpha \boldsymbol{z} - \boldsymbol{G}(E_n^\alpha)\|_{\ell_\infty} = 0.$$

This in particular implies that (note that limits here are with respect to the natural norm of $\mathbb{X}$)

$$\lim_{n\to\infty} \|\alpha \boldsymbol{G}_{\boldsymbol{p}}(\mathcal{H}) - \boldsymbol{G}_{\boldsymbol{p}}(E_n^\alpha)\|_{\ell_\infty} = 0,$$

and by a symmetric argument,

$$\lim_{n\to\infty} \|(1-\alpha)\boldsymbol{G}_{\boldsymbol{p}'}(\mathcal{H}) - \boldsymbol{G}_{\boldsymbol{p}'}((E_n^\alpha)^c)\|_{\ell_\infty} = 0.$$

Now, we may define the sequence of policies

$$\boldsymbol{p}_\alpha^n(\boldsymbol{h}) = \mathbb{1}_{E_n^\alpha}(\boldsymbol{h})\boldsymbol{p}(\boldsymbol{h}) + \mathbb{1}_{\mathcal{H}\backslash E_n^\alpha}(\boldsymbol{h})\boldsymbol{p}'(\boldsymbol{h}) = \begin{cases} \boldsymbol{p}(\boldsymbol{h}), & \text{if } \boldsymbol{h} \in E_n^\alpha \\ \boldsymbol{p}'(\boldsymbol{h}), & \text{if } \boldsymbol{h} \in \mathcal{H}\backslash E_n^\alpha \end{cases}, \quad n \in \mathbb{N}.$$

Of course, it holds that $\boldsymbol{p}_\alpha^n \in \Pi$ for all $n \in \mathbb{N}$ because Π is decomposable. Then, it follows that

$$\begin{aligned} &\|\boldsymbol{G}_{\boldsymbol{p}_\alpha^n}(\mathcal{H}) - \alpha \boldsymbol{G}_{\boldsymbol{p}}(\mathcal{H}) - (1-\alpha)\boldsymbol{G}_{\boldsymbol{p}'}(\mathcal{H})\|_{\ell_\infty} \\ &\quad = \|\boldsymbol{G}_{\boldsymbol{p}_\alpha^n}(E_n^\alpha) + \boldsymbol{G}_{\boldsymbol{p}_\alpha^n}((E_n^\alpha)^c) - \alpha \boldsymbol{G}_{\boldsymbol{p}}(\mathcal{H}) - (1-\alpha)\boldsymbol{G}_{\boldsymbol{p}'}(\mathcal{H})\|_{\ell_\infty} \\ &\quad = \|\boldsymbol{G}_{\boldsymbol{p}}(E_n^\alpha) + \boldsymbol{G}_{\boldsymbol{p}'}((E_n^\alpha)^c) - \alpha \boldsymbol{G}_{\boldsymbol{p}}(\mathcal{H}) - (1-\alpha)\boldsymbol{G}_{\boldsymbol{p}'}(\mathcal{H})\|_{\ell_\infty} \\ &\quad \le \|\boldsymbol{G}_{\boldsymbol{p}}(E_n^\alpha) - \alpha \boldsymbol{G}_{\boldsymbol{p}}(\mathcal{H})\|_{\ell_\infty} + \|\boldsymbol{G}_{\boldsymbol{p}'}((E_n^\alpha)^c) - (1-\alpha)\boldsymbol{G}_{\boldsymbol{p}'}(\mathcal{H})\|_{\ell_\infty}, \end{aligned}$$

which implies that

$$\lim_{n\to\infty} \|\boldsymbol{G}_{\boldsymbol{p}_\alpha^n}(\mathcal{H}) - \alpha \boldsymbol{G}_{\boldsymbol{p}}(\mathcal{H}) - (1-\alpha)\boldsymbol{G}_{\boldsymbol{p}'}(\mathcal{H})\|_{\ell_\infty} = 0.$$

Equivalently, we have shown that for every $\varepsilon > 0$, there exists a positive number $N(\varepsilon) > 0$, such that for every $n > N(\varepsilon)$,

$$\begin{aligned} &\|\mathbb{E}\{\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{H}) \otimes \boldsymbol{f}(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})\} \\ &\quad - \alpha\mathbb{E}\{\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{H}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} - (1-\alpha)\mathbb{E}\{\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{H}) \otimes \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})\}\|_{\ell_\infty} \le \varepsilon. \end{aligned}$$

Evidently, $N(\varepsilon)$ is uniform over the individual elements of the countable basis $\mathbb{B}$. We can rewrite the preceding expression as

$$\begin{aligned} &|\mathbb{E}\{\lambda_m(\boldsymbol{H}) f_i(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})\} \\ &\quad - \alpha\mathbb{E}\{\lambda_m(\boldsymbol{H}) f_i(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} - (1-\alpha)\mathbb{E}\{\lambda_m(\boldsymbol{H}) f_i(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})\}| \le \varepsilon, \end{aligned}$$

for every pair $(m,i) \in \mathbb{N} \times \mathbb{N}_N^+$. Now, for each choice of $\zeta \in \mathbb{A}_\gamma^i$, $i \in \mathbb{N}_N^+$, we can extract a subsequence $\{\lambda_m\}_{m\in\mathcal{K}}, \mathcal{K} \subseteq \mathbb{N}$ converging to ζ in $\mathcal{L}_1$. A consequence of this is the existence of a further sub-subsequence $\{\lambda_m\}_{m\in\mathcal{K}'}$, $\mathcal{K}' \subseteq \mathcal{K}$ such that

$$\lambda_m \underset{\mathcal{K}' \ni m\to\infty}{\longrightarrow} \zeta, \quad \text{P-a.e.}$$

Then, by dominated convergence, we have for each i-th element of the service vector $\boldsymbol{f}$,

$$\begin{aligned} \mathbb{E}\{\lambda_m(\boldsymbol{H}) f_i(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})\} &\underset{\mathcal{K}' \ni m\to\infty}{\longrightarrow} \mathbb{E}\{\zeta(\boldsymbol{H}) f_i(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})\}, \\ \mathbb{E}\{\lambda_m(\boldsymbol{H}) f_i(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} &\underset{\mathcal{K}' \ni m\to\infty}{\longrightarrow} \mathbb{E}\{\zeta(\boldsymbol{H}) f_i(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} \quad \text{and} \\ \mathbb{E}\{\lambda_m(\boldsymbol{H}) f_i(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})\} &\underset{\mathcal{K}' \ni m\to\infty}{\longrightarrow} \mathbb{E}\{\zeta(\boldsymbol{H}) f_i(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})\}. \end{aligned}$$

Therefore, we have that, for every $\varepsilon > 0$ there exists a positive number $N(\varepsilon) > 0$, such that for every $n > N(\varepsilon)$ and for every $\zeta \in \mathbb{A}_\gamma^i$,[2]

$$\begin{aligned}&|\mathbb{E}\{\zeta(\boldsymbol{H}) f_i(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})\} \\ &\quad - \alpha\mathbb{E}\{\zeta(\boldsymbol{H}) f_i(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} - (1-\alpha)\mathbb{E}\{\zeta(\boldsymbol{H}) f_i(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})\}| \leq \varepsilon, \quad i \in \mathbb{N}_N^+.\end{aligned}$$

The last expression implies in particular that

$$\begin{aligned}&\Bigg| \inf_{\zeta \in \mathbb{A}_\gamma^i} \mathbb{E}\{\zeta(\boldsymbol{H}) f_i(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})\} \\ &\quad - \inf_{\zeta \in \mathbb{A}_\gamma^i} \mathbb{E}\{\zeta(\boldsymbol{H})[\alpha f_i(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) + (1-\alpha) f_i(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})]\}\Bigg| \leq \varepsilon, \quad i \in \mathbb{N}_N^+,\end{aligned}$$

which is the same as

$$\left\| \inf_{\boldsymbol{\zeta} \in \mathbb{A}_\gamma^S} \mathbb{E}\{\boldsymbol{\zeta}(\boldsymbol{H}) \circ \boldsymbol{f}(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})\} - \inf_{\boldsymbol{\zeta} \in \mathbb{A}_\gamma^S} \mathbb{E}\{\boldsymbol{\zeta}(\boldsymbol{H}) \circ [\alpha \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) + (1-\alpha) \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})]\} \right\|_\infty \leq \varepsilon.$$

By risk duality, we obtain that, for every $\varepsilon > 0$, there is $N(\varepsilon) > 0$ such that for every $n > N(\varepsilon)$, it is true that

$$\| -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H})) + \boldsymbol{\rho}(-\alpha \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) - (1-\alpha) \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H}))\|_\infty \leq \varepsilon.$$

This implies that, for every choice of $\boldsymbol{p} \in \Pi$, $\boldsymbol{p}' \in \Pi$, for every $\alpha \in [0,1]$ and for every $\varepsilon > 0$, there exists at least one policy $\boldsymbol{p}_\alpha^\varepsilon \in \Pi$ (in fact, a whole family of such policies) such that

$$-\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha^\varepsilon(\boldsymbol{H}), \boldsymbol{H})) + \boldsymbol{\rho}(-\alpha \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) - (1-\alpha) \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})) \geq -\varepsilon \mathbf{1}.$$

This fact together with convexity of $\boldsymbol{\rho}$ further implies that

$$\begin{aligned}-\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha^\varepsilon(\boldsymbol{H}), \boldsymbol{H})) &\geq -\boldsymbol{\rho}(-\alpha \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) - (1-\alpha) \boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})) - \varepsilon \mathbf{1} \\ &= -\alpha \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) - (1-\alpha) \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})) - \varepsilon \mathbf{1}.\end{aligned}$$

Let us now see how the preceding fact implies that $\mathrm{cl}(\mathcal{C})$ is a convex set. Let $(\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \in \mathcal{C}$ for $(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi$ and let $(\delta_o', \boldsymbol{\delta}_r', \boldsymbol{\delta}_d') \in \mathcal{C}$ for $(\boldsymbol{x}', \boldsymbol{p}') \in \mathcal{X} \times \Pi$. Also choose a sequence $\{\varepsilon_n > 0\}_{n \in \mathbb{N}}$ decreasing to zero, say $\varepsilon_n \triangleq 1/(n+1), n \in \mathbb{N}$. By our discussion above, we have actually shown that, for every $\alpha \in [0,1]$, it holds that, for every n,

$$\alpha(\delta_o, \boldsymbol{\delta}_r - \varepsilon_n \mathbf{1}, \boldsymbol{\delta}_d) + (1-\alpha)(\delta_o', \boldsymbol{\delta}_r' - \varepsilon_n \mathbf{1}, \boldsymbol{\delta}_d') \in \mathcal{C}.$$

To verify this claim, observe that, for every $n \in \mathbb{N}$, there exists a policy $\boldsymbol{p}_\alpha^{\varepsilon_n} \in \Pi$ such that for the pair $(\boldsymbol{x}_\alpha = \alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{x}', \boldsymbol{p}_\alpha^{\varepsilon_n}) \in \mathcal{X} \times \Pi$ (by assumption, $\mathcal{X}$ is convex) it is true that

$$\begin{aligned}&-\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha^{\varepsilon_n}(\boldsymbol{H}), \boldsymbol{H})) - \boldsymbol{x}_\alpha \\ &\quad = -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha^{\varepsilon_n}(\boldsymbol{H}), \boldsymbol{H})) - \alpha \boldsymbol{x} - (1-\alpha)\boldsymbol{x}' \\ &\quad \geq -\alpha \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) - (1-\alpha) \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})) - \alpha \boldsymbol{x} - (1-\alpha)\boldsymbol{x}' - \varepsilon_n \mathbf{1} \\ &\quad = -\alpha \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) - \alpha \boldsymbol{x} - (1-\alpha) \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})) - (1-\alpha)\boldsymbol{x}' - \varepsilon_n \mathbf{1} \\ &\quad \geq \alpha \boldsymbol{\delta}_r + (1-\alpha) \boldsymbol{\delta}_r' - \varepsilon_n \mathbf{1},\end{aligned}$$

[2] Note that the basic fact that enables interchanging the order of limits relative to n and m above is that convergence over n is uniform over m, i.e., the index of the elements in the dense set $\mathbb{B}$. Then, we reiterate the same procedure for every ζ in each of the risk envelopes $\mathbb{A}_\gamma^i, i \in \mathbb{N}_N^+$, by extracting a different subsequence out of $\mathbb{B}$ each time.

or, equivalently,

$$-\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha^{\varepsilon_n}(\boldsymbol{H}),\boldsymbol{H}))-\boldsymbol{x}_\alpha \geq \alpha(\boldsymbol{\delta}_r-\varepsilon_n\mathbf{1})+(1-\alpha)(\boldsymbol{\delta}_r'-\varepsilon_n\mathbf{1}).$$

This means that the pair $(\boldsymbol{x}_\alpha,\boldsymbol{p}_\alpha^{\varepsilon_n})\in\mathcal{X}\times\Pi$ is such that

$$\begin{gathered} g^o(\boldsymbol{x}_\alpha)\geq\alpha\delta_o+(1-\alpha)\delta_o', \\ -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}_\alpha^{\varepsilon_n}(\boldsymbol{H}),\boldsymbol{H}))-\boldsymbol{x}_\alpha \geq \alpha(\boldsymbol{\delta}_r-\varepsilon_n\mathbf{1})+(1-\alpha)(\boldsymbol{\delta}_r'-\varepsilon_n\mathbf{1}) \quad \text{and} \\ \boldsymbol{g}(\boldsymbol{x}_\alpha)\geq\alpha\boldsymbol{\delta}_d+(1-\alpha)\boldsymbol{\delta}_d', \end{gathered}$$

verifying our claim. But, evidently,

$$\lim_{n\to\infty}\alpha(\delta_o,\boldsymbol{\delta}_r-\varepsilon_n\mathbf{1},\boldsymbol{\delta}_d)+(1-\alpha)(\delta_o',\boldsymbol{\delta}_r'-\varepsilon_n\mathbf{1},\boldsymbol{\delta}_d')=\alpha(\delta_o,\boldsymbol{\delta}_r,\boldsymbol{\delta}_d)+(1-\alpha)(\delta_o',\boldsymbol{\delta}_r',\boldsymbol{\delta}_d'),$$

which of course implies that

$$\alpha(\delta_o,\boldsymbol{\delta}_r,\boldsymbol{\delta}_d)+(1-\alpha)(\delta_o',\boldsymbol{\delta}_r',\boldsymbol{\delta}_d')\in\mathrm{cl}(\mathcal{C}),$$

as a limit point of elements in $\mathcal{C}$. Finally, convexity of $\mathrm{cl}(\mathcal{C})$ follows in light of the next elementary result, whose proof we also present for completeness.

Proposition 4. *Suppose that a set $\mathcal{A}\subseteq\mathbb{R}^N$ has the property that all convex combinations of any two points in $\mathcal{A}$ belong to its closure, i.e.,*

$$\boldsymbol{x}\in\mathcal{A} \quad \textit{and} \quad \boldsymbol{y}\in\mathcal{A} \implies \forall\alpha\in[0,1], \quad \alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}\in\mathrm{cl}(\mathcal{A}).$$

Then $\mathrm{cl}(\mathcal{A})$ *is (closed) convex.*

Proof of Proposition 4. Let $\boldsymbol{x}\in\mathrm{cl}(\mathcal{A})$ and $\boldsymbol{y}\in\mathrm{cl}(\mathcal{A})$. Then, we can find sequences $\{\boldsymbol{x}_n\in\mathcal{A}\}_{n\in\mathbb{N}}$ and $\{\boldsymbol{y}_n\in\mathcal{A}\}_{n\in\mathbb{N}}$ such that

$$\boldsymbol{x}=\lim_{n\to\infty}\boldsymbol{x}_n \quad \text{and} \quad \boldsymbol{y}=\lim_{n\to\infty}\boldsymbol{y}_n,$$

where the limits are interpreted in the standard Euclidean sense. So, given $\alpha\in[0,1]$,

$$\begin{aligned} \alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y} &= \alpha\lim_{n\to\infty}\boldsymbol{x}_n+(1-\alpha)\lim_{n\to\infty}\boldsymbol{y}_n \\ &= \lim_{n\to\infty}\alpha\boldsymbol{x}_n+(1-\alpha)\boldsymbol{y}_n. \end{aligned}$$

However, by assumption for the set $\mathcal{A}$ it holds that, for every $n\in\mathbb{N}$,

$$\alpha\boldsymbol{x}_n+(1-\alpha)\boldsymbol{y}_n\in\mathrm{cl}(\mathcal{A}),$$

and the sequence converges; therefore it must converge in $\mathrm{cl}(\mathcal{A})$, and the limit is $\alpha\boldsymbol{x}+(1-\alpha)\boldsymbol{y}$. Therefore, $\mathrm{cl}(\mathcal{A})$ is closed convex. ∎

By a trivial application of Proposition 4, we obtain that $\mathrm{cl}(\mathcal{C})$ is a (closed) convex set, and the proof is now complete. ∎

5.3 Convexity of cl($\mathcal{C}$) Implies Strong Duality

Let us now finish the proof of Theorem 1 by exploiting the convexity of the closure of the utility-constraint set

$$\mathcal{C}=\left\{(\delta_o,\boldsymbol{\delta}_r,\boldsymbol{\delta}_d)\,\middle|\,\begin{array}{l} g^o(\boldsymbol{x})\geq\delta_o \\ -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H}))-\boldsymbol{x}\geq\boldsymbol{\delta}_r \text{ for some } (\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi \\ \boldsymbol{g}(\boldsymbol{x})\geq\boldsymbol{\delta}_d \end{array}\right\},$$

the expression of which we repeat here for convenience, together with condition (5) of Assumption 1, namely that problem (RCP) satisfies Slater's condition. Let us recall the Lagrangian associated with problem (RCP), i.e.,

$$\mathsf{L}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\lambda}) = g^o(\boldsymbol{x}) + \langle \boldsymbol{\lambda_g}, \boldsymbol{g}(\boldsymbol{x})\rangle + \big\langle \boldsymbol{\lambda_\rho}, -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) - \boldsymbol{x}\big\rangle,$$

where $(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\lambda}) \in \mathbf{C} \times \Pi \times \mathbb{R}^{N_g} \times \mathbf{C}$, for which we already know that

$$\inf_{\boldsymbol{\lambda}\geq \mathbf{0}} \sup_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi} \mathsf{L}(\boldsymbol{x},\boldsymbol{p},\boldsymbol{\lambda}) = \mathsf{D}^* \geq \mathsf{P}^*.$$

Strong duality of (RCP) will be immediate if we can show that $\mathsf{D}^* \leq \mathsf{P}^*$.

Our discussion substantially extends [22, Theorem 1 and its proof] and [23, Appendix A], and is eventually based on a standard application of the supporting hyperplane theorem (see e.g., [5, Proposition 1.5.1]), which we outline below for completeness. Note that the same technique would be applicable in case our initial problem (RCP) was originally convex. The reason is that the only fact needed at this point is the "top-level" convexity of the $\mathrm{cl}(\mathcal{C})$, which would come for free if (RCP) was itself a convex program. Before we assemble everything together, we need an additional technical result.

Lemma 5 (**Point on the Shell**). *$(\mathsf{P}^*,\mathbf{0},\mathbf{0})$ is not in the interior of $\mathrm{cl}(\mathcal{C})$.*

Proof of Lemma 5. First, we observe that the point $(\mathsf{P}^*,\mathbf{0},\mathbf{0})$ cannot be in the interior of $\mathcal{C}$, for otherwise there would exist an $\varepsilon > 0$ such that $(\mathsf{P}^*+\varepsilon,\mathbf{0},\mathbf{0}) \in \mathcal{C}$, contradicting P^* being the optimal value of the initial constrained problem. In fact, if $(\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_d^+) > \mathbf{0}$, every perturbation of the form

$$(\mathsf{P}^*,\mathbf{0},\mathbf{0}) + (\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_d^+) = (\mathsf{P}^*+\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_r^+)$$

cannot be in $\mathcal{C}$ either; if it was, this would imply the existence of a pair $(\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X}\times\Pi$ attaining a strictly larger objective than P^*, while also keeping all the constraints inactive. Therefore, it follows that the *open orthant (an open convex set)*

$$\mathcal{D} = \{(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) | (\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = (\mathsf{P}^*+\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_r^+), \quad (\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_r^+) > \mathbf{0}\}$$

is disjoint from $\mathcal{C}$.

Let us show that, in fact, $\mathcal{D}$ and $\mathrm{cl}(\mathcal{C})$ are also disjoint. Note that $\mathcal{D}$ has a nonempty interior in the standard Euclidean topology of $\mathcal{C}$. Therefore, for every element of $\mathcal{D}$ there exists an open ball of the same dimension as that of $\mathcal{C}$ entirely contained in $\mathcal{D}$ and having that element of $\mathcal{D}$ as its center. Next, if an element of $\mathcal{D}$, say $\boldsymbol{d}$, was in $\mathrm{cl}(\mathcal{C})$, there should exist a sequence contained entirely in $\mathcal{C}$ converging to $\boldsymbol{d}$. However, since $\boldsymbol{d}$ is contained in an open ball, say $\mathcal{B}_{\boldsymbol{d}}$, entirely contained in $\mathcal{D}$ and of the same dimension as that of $\mathcal{C}$, every sequence in $\mathcal{C}$ converging to $\boldsymbol{d}$ must break into that open ball. This implies that the aforementioned sequence in $\mathcal{C}$ converging to $\boldsymbol{d}$ must have elements also in $\mathcal{B}_{\boldsymbol{d}}$. This is absurd, since $\mathcal{C}\bigcap\mathcal{B}_{\boldsymbol{d}} = \emptyset$ (due to the fact that $\mathcal{C}\bigcap\mathcal{D} = \emptyset$).

Now suppose that $(\mathsf{P}^*,\mathbf{0},\mathbf{0})$ is in the interior of $\mathrm{cl}(\mathcal{C})$. Then, there must exist another open ball, say $\mathcal{B}_*$, of the same dimension as that of $\mathcal{C}$, completely contained in $\mathrm{cl}(\mathcal{C})$ and centered at $(\mathsf{P}^*,\mathbf{0},\mathbf{0})$. This means that there exists a positive perturbation vector $(\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_r^+) > \mathbf{0}$ such that the point $(\mathsf{P}^*+\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_r^+)$ is in $\mathcal{B}_*$. This is also absurd, because $(\mathsf{P}^*+\delta_o^+,\boldsymbol{\delta}_r^+,\boldsymbol{\delta}_r^+) \in \mathcal{D}$ and $\mathcal{D}\bigcap\mathrm{cl}(\mathcal{C}) = \emptyset$, as shown above. ■

The complete argument based on a standard application of the supporting hyperplane theorem now follows.

Theorem 6 (**Supporting Hyperplane Theorem**). *Let $\mathcal{A} \subseteq \mathbb{R}^n$ be a nonempty convex set. If $\boldsymbol{\delta}^* \in \mathbb{R}^n$ is not in the interior of $\mathcal{A}$, then there exists a hyperplane passing through $\boldsymbol{\delta}^*$ such that $\mathcal{A}$ is in one of its closed halfspaces. In other words, there exists a vector $\boldsymbol{\lambda} \neq \mathbf{0}$ such that, for every $\boldsymbol{\delta} \in \mathcal{A}$, it holds that $\langle \boldsymbol{\lambda}, \boldsymbol{\delta}^*\rangle \geq \langle \boldsymbol{\lambda}, \boldsymbol{\delta}\rangle$.*

Remark 1. *Note that the inequality in the supporting hyperplane theorem can be equivalently reversed by flipping the sign of the support vector* $\boldsymbol{\lambda}$.

We start by observing that since (RCP) satisfies Slater's condition, it follows that $\mathcal{C}$ is nonempty, and the same follows for $\mathrm{cl}(\mathcal{C})$. So $\mathrm{cl}(\mathcal{C})$ is a nonempty (closed) convex set. By Lemma 5, we also know that the point $(\mathsf{P}^*, \mathbf{0}, \mathbf{0})$ is not in the interior of $\mathrm{cl}(\mathcal{C})$. We can then apply the supporting hyperplane theorem on the pair $\mathrm{cl}(\mathcal{C})$ and $(\mathsf{P}^*, \mathbf{0}, \mathbf{0})$, implying existence of a vector of multipliers $(\lambda_o, \boldsymbol{\lambda_g}, \boldsymbol{\lambda_\rho}) \neq \mathbf{0}$ such that, for every $(\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \in \mathrm{cl}(\mathcal{C})$, it is true that

$$\lambda_o \delta_o + \langle \boldsymbol{\lambda_g}, \boldsymbol{\delta}_r \rangle + \langle \boldsymbol{\lambda_\rho}, \boldsymbol{\delta}_d \rangle \leq \lambda_o \mathsf{P}^*.$$

It readily follows that, in fact, $(\lambda_o, \boldsymbol{\lambda_g}, \boldsymbol{\lambda_\rho}) \geq \mathbf{0}$. Indeed, if any component of $(\lambda_o, \boldsymbol{\lambda_g}, \boldsymbol{\lambda_\rho})$ was negative, then we could choose $(\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \in \mathcal{C} \subseteq \mathrm{cl}(\mathcal{C})$ such that the corresponding inner product becomes arbitrarily large (note that $\mathcal{C}$ is unbounded below), eventually violating the inequality above, regardless of the sign of P^*.

The second fact we may show is that $\lambda_o \neq 0$, which implies that $\lambda_o > 0$. Again, if $\lambda_o = 0$, then we would have that

$$\langle \boldsymbol{\lambda_g}, \boldsymbol{\delta}_r \rangle + \langle \boldsymbol{\lambda_\rho}, \boldsymbol{\delta}_d \rangle \leq 0.$$

But $(\boldsymbol{\lambda_g}, \boldsymbol{\lambda_\rho}) \neq \mathbf{0}$ (i.e., there is at least one nonzero component) and $(\boldsymbol{\lambda_g}, \boldsymbol{\lambda_\rho}) \geq \mathbf{0}$ (shown above), and problem (RCP) satisfies Slater's condition, so the preceding inequality is also absurd.

As a result, we may divide by $\lambda_o > 0$, showing that there is $(\boldsymbol{\lambda}^\star_{\boldsymbol{g}} \triangleq \boldsymbol{\lambda_g}/\lambda_o, \boldsymbol{\lambda}^\star_{\boldsymbol{\rho}} \triangleq \boldsymbol{\lambda_\rho}/\lambda_o) \geq \mathbf{0}$, such that, for every $(\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \in \mathrm{cl}(\mathcal{C})$, it holds that

$$\delta_o + \langle \boldsymbol{\lambda}^\star_{\boldsymbol{g}}, \boldsymbol{\delta}_r \rangle + \langle \boldsymbol{\lambda}^\star_{\boldsymbol{\rho}}, \boldsymbol{\delta}_d \rangle \leq \mathsf{P}^*.$$

By construction of the set $\mathrm{cl}(\mathcal{C})$ (as the closure of $\mathcal{C}$), we obtain in particular that

$$\mathsf{L}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}^\star) = g^o(\boldsymbol{x}) + \langle \boldsymbol{\lambda}^\star_{\boldsymbol{g}}, \boldsymbol{g}(\boldsymbol{x}) \rangle + \left\langle \boldsymbol{\lambda}^\star_{\boldsymbol{\rho}}, -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) - \boldsymbol{x} \right\rangle \leq \mathsf{P}^*,$$

for every pair $(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi$, since all such pairs correspond to points included in $\mathcal{C} \subseteq \mathrm{cl}(\mathcal{C})$. This further implies that we are allowed to maximize both sides of the inequality over all possible $(\boldsymbol{x}, \boldsymbol{p})$, yielding

$$\begin{aligned} -\infty < \mathsf{D}^* &= \inf_{\boldsymbol{\lambda} \geq \mathbf{0}} \sup_{(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi} \mathsf{L}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}) \\ &\leq \sup_{(\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi} \mathsf{L}(\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{\lambda}^\star) \leq \mathsf{P}^*, \end{aligned}$$

and we are done. ■

6 Extensions and Implications

6.1 Model Generalizations

Let us now briefly demonstrate how our main result in Theorem 1 can be applied to certain, possibly less obvious variations or generalizations of the base model in (RCP).

Firstly, it readily follows that strong duality of problem (RCP) implies strong duality for problems in which risk components also appear in the objective. To see this, consider the simplistic problem

$$\infty > \mathsf{P}^* = \begin{array}{rl} \underset{\boldsymbol{x}, \boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) - \rho(-f(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) \\ \text{subject to} & \boldsymbol{g}(\boldsymbol{x}) \geq \mathbf{0} \\ & (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array}, \tag{RCP$'$}$$

where $\rho : \mathcal{L}_1(\mathrm{P}, \mathbb{R}) \to \mathbb{R}$ is a convex, lower semicontinuous, and positively homogeneous risk measure, and $f : \mathbf{R} \times \mathcal{H} \to \mathbb{R}$ is an arbitrary cost function such that $f(\cdot, \boldsymbol{H}) \in \mathcal{L}_1(\mathrm{P}, \mathbb{R})$. Evidently, the value of (RCP′), P^*, coincides with that of the problem

$$\begin{aligned} \bar{\mathsf{P}}^* \triangleq \underset{\boldsymbol{x},t,\boldsymbol{p}(\cdot)}{\text{maximize}} \;& g^o(\boldsymbol{x}) + t \\ \text{subject to} \;& t \le -\rho(-f(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) \\ & \boldsymbol{g}(\boldsymbol{x}) \ge \mathbf{0} \\ & (\boldsymbol{x},t,\boldsymbol{p}) \in \mathcal{X}\times\mathbb{R}\times\Pi \end{aligned},$$

and then it readily follows that (RCP′) exhibits strong duality, assuming it satisfies Assumption 1. To see this, we may define the corresponding Lagrangian as

$$\bar{\mathsf{L}}(\boldsymbol{x},t,\boldsymbol{p},\boldsymbol{\lambda}) \triangleq g^o(\boldsymbol{x}) + t + \langle \boldsymbol{\lambda}_{\boldsymbol{g}}, \boldsymbol{g}(\boldsymbol{x})\rangle + \lambda_\rho\left(-\rho(-f(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) - t\right),$$

where $\boldsymbol{\lambda} \equiv (\boldsymbol{\lambda}_{\boldsymbol{g}}, \lambda_\rho) \in \mathbb{R}^{N_{\boldsymbol{g}}} \times \mathbb{R}$ are the multipliers associated with the dualized constraints. Then, from Theorem 1, we obtain that

$$\begin{aligned} \mathsf{P}^* \equiv \bar{\mathsf{P}}^* = \bar{\mathsf{D}}^* &\triangleq \inf_{\boldsymbol{\lambda}\ge\mathbf{0}} \sup_{(\boldsymbol{x},t,\boldsymbol{p})\in\mathcal{X}\times\Pi} \bar{\mathsf{L}}(\boldsymbol{x},t,\boldsymbol{p},\boldsymbol{\lambda}) \\ &= \inf_{\boldsymbol{\lambda}\ge\mathbf{0}} \sup_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi} g^o(\boldsymbol{x}) - \lambda_\rho\rho(-f(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) + \langle \boldsymbol{\lambda}_{\boldsymbol{g}}, \boldsymbol{g}(\boldsymbol{x})\rangle + \sup_{t\in\mathbb{R}} t(1-\lambda_\rho) \\ &= \inf_{\boldsymbol{\lambda}_{\boldsymbol{g}}\ge\mathbf{0}} \sup_{(\boldsymbol{x},\boldsymbol{p})\in\mathcal{X}\times\Pi} g^o(\boldsymbol{x}) - \rho(-f(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) + \langle \boldsymbol{\lambda}_{\boldsymbol{g}}, \boldsymbol{g}(\boldsymbol{x})\rangle = \mathsf{D}^*, \end{aligned}$$

where by inspection we observe that the choice $\lambda_\rho = 1$ is optimal (since otherwise $\bar{\mathsf{D}}^* = \infty$). Hence, we verify that $\mathsf{P}^* = \mathsf{D}^*$ and, in fact, (RCP′) must necessarily exhibit strong duality.

Another, perhaps more trivial, extension of our base risk-constrained setting would be qualifying problems of the form

$$\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} \;& g^o(\boldsymbol{x}) \\ \text{subject to} \;& \boldsymbol{w}(\boldsymbol{x}) \le -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) \\ & \boldsymbol{g}(\boldsymbol{x}) \ge \mathbf{0} \\ & (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X}\times\Pi \end{aligned},$$

where $\boldsymbol{w}$ is a convex function. Indeed, the proof would follow exactly the developments of Section 5, requiring only certain trivial minor modifications. This was omitted for simplicity, since (RCP) was inspired by well-known resource allocation problems, where $\boldsymbol{w}(\boldsymbol{x}) = \boldsymbol{x}$; see, e.g., Sections 1.1 and 7.

Putting it altogether, we see that Theorem 1 is applicable to general *risk-over-risk problems* of the form

$$\boxed{\begin{aligned} \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} \;& g^o(\boldsymbol{x}) + \sum_{j=1}^{N_o} -\rho_j^o(-f_j^o(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) \\ \text{subject to} \;& \boldsymbol{w}(\boldsymbol{x}) \le -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) \\ & \boldsymbol{g}(\boldsymbol{x}) \ge \mathbf{0} \\ & (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X}\times\Pi \end{aligned},} \qquad \text{(ROR)}$$

under the appropriate structural conditions on the involved risk measures and service functions, as those are implied by the discussion above.

Lastly, we conjecture that Theorem 1 can be established for more general conic formulations of (RCP) (that is, replacing inequality constraints by convex conic inequalities), by properly extending our proof, for instance utilizing results developed in [30]; see also Theorem 8 discussed later in Appendix C. This more technical treatment is omitted here, for simplicity of exposition.

6.2 CVaRizations and the Case of Risk Measures on $\mathcal{L}_p, p > 1$

The duality theory developed so far does *not* readily extend to real-valued convex, lower semicontinuous and positively homogeneous risk measures on the space $\mathcal{L}_p(\mathrm{P},\mathbb{R})$, for $p > 1$. Thus, it remains an open problem to show that Theorem 1 or some version of it holds for such risk measures. While we leave a complete answer to this question for future work, we conjecture that such an extension of our main result would probably need to rely on another construction, or different result from Uhl's theorem (Theorem 2); one possibility could be the strong Lyapunov theorem for the weak topology, due to Knowles (see [40], also [29, Theorem IX.1.4]).

Still, due to the fact that many useful risk measures are naturally defined on favorable subsets of $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$ –such as the reflexive spaces $\mathcal{L}_p(\mathrm{P}, \mathbb{R})$, $p \in (1, \infty)$; a classical example is the mean-upper-semideviation of order p [1, Example 6.23]–, it remains interesting to investigate ways to approximate a risk measure on $\mathcal{L}_p(\mathrm{P}, \mathbb{R}), p > 1$, via another related risk measure taking finite values on $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$, so that Theorem 1 can still be applied by relying on this approximation. One general way to do this is via a CVaR*ization* operation. In particular, given some proper, lower semicontinuous functional ρ on $\mathcal{L}_p(\mathrm{P}, \mathbb{R})$, $p \in [1, \infty)$, we define its CVaR *envelope* at level $\beta \in (0, 1]$, denoted by $\rho_\beta \colon \mathcal{L}_1(\mathrm{P}, \mathbb{R}) \to \mathbb{R} \cup \{\pm\infty\} \triangleq \overline{\mathbb{R}}$, as

$$\rho_\beta(Z) \equiv (\rho \,\square\, \mathrm{CVaR}^\beta)(Z) \triangleq \inf_{Y \in \mathcal{L}_1(\mathrm{P}, \mathbb{R})} \{\rho\,(Z - Y) + \mathrm{CVaR}^\beta(Y)\},$$

where we are dropping dependencies on $\boldsymbol{H}$ throughout this section for brevity of exposition. Despite ρ being finite-valued only on $\mathcal{L}_p(\mathrm{P}, \mathbb{R}) \subset \mathcal{L}_1(\mathrm{P}, \mathbb{R})$, the infimal convolution of ρ with the CVaR ensures that the domain of the resulting envelope is the whole space $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$ (indeed, see [45, Theorem 2.2]). Let us also define the *meet* of two extended real-valued functionals ρ_1 and ρ_2 as $(\rho_1 \wedge \rho_2)(Z) = \min\{\rho_1(Z), \rho_2(Z)\}$. In what follows, we invoke the following blanket assumption.

Assumption 2. *The functional $\rho \colon \mathcal{L}_p(\mathrm{P}, \mathbb{R}) \to \overline{\mathbb{R}}$, $p \in [1, \infty)$, is proper, convex, and lower semi-continuous. Further, for every $\beta \in [0, 1]$, $\operatorname{dom} \rho^* - \mathbb{A}_\beta$ is a neighbourhood of the origin, where $\mathbb{A}_\beta$ is the risk envelope of CVaR^β, with $\mathrm{CVaR}^0 \triangleq \operatorname{ess\,sup}$.*

Let us note that the above regularity condition (i.e. Assumption 2) is standard (e.g. see [46, Chapter 9]), and it can further be relaxed, if necessary, as shown in [47].

Theorem 7 (**CVaRizations**)**.** *Let ρ be a risk measure satisfying Assumption 2, and consider its CVaR envelope $\rho_\beta \equiv \rho \,\square\, \mathrm{CVaR}^\beta \colon \mathcal{L}_r(\mathrm{P}, \mathbb{R}) \to \mathbb{R}$, for $r \in [1, p]$ and $\beta \in (0, 1]$. The following hold:*

1. *If $r > 1$, ρ_β is exact (i.e., the infimum is attained).*
2. *$\inf \rho_\beta = \inf \rho + \inf \mathrm{CVaR}^\beta$. Additionally, $\arg\min \rho_\beta \supseteq \arg\min \rho + \arg\min \mathrm{CVaR}^\beta$, with equality if $\inf \rho$ is a real number and ρ_β is exact.*
3. *If ρ is subadditive, then the envelope ρ_β is subadditive on $\mathcal{L}_1$, and it is the largest subadditive minorant of $\rho \wedge \mathrm{CVaR}^\beta$. In fact, if the meet is subadditive, then $\rho_\beta = \rho \wedge \mathrm{CVaR}^\beta$.*
4. *The envelope ρ_β is convex continuous on $\mathcal{L}_1(P, \mathbb{R})$ (thus subdifferentiable), and admits*

$$(\rho \,\square\, \mathrm{CVaR}^\beta)(Z) = \sup_{\zeta \in \mathbb{A}_\beta} \{\langle \zeta, Z \rangle - \rho^*(\zeta)\},$$
$$\partial(\rho \,\square\, \mathrm{CVaR}^\beta)(Z) = \operatorname*{arg\,max}_{\zeta \in \mathbb{A}_\beta} \{\langle \zeta, Z \rangle - \rho^*(\zeta)\}.$$

   Further, ρ_β is monotone and translation equivariant, even if ρ is not.

5. *If ρ is positively homogeneous, then ρ_β coherent. In this case, $\rho_\beta = (\rho_\beta^*)^*$ and*

$$\rho_\beta^* = \rho^* + \mathrm{CVaR}_\beta^* \equiv \delta_{\widehat{\mathbb{A}}_\beta},$$

   where $\widehat{\mathbb{A}}_\beta = \{\zeta \in \mathbb{A} \big| \zeta \leq 1/\beta, \zeta \geq 0, \mathbb{E}\{\zeta\} = 1\}$ and $\delta_{\widehat{\mathbb{A}}_\beta}$ is the indicator to $\widehat{\mathbb{A}}_\beta$. Further, if ρ is itself coherent, then $\widehat{\mathbb{A}}_\beta = \{\zeta \in \mathbb{A} \big| \zeta \leq 1/\beta\}$, and letting $r = p > 1$, ρ_β converges in the Mosco sense[3] to ρ, as $\beta \searrow 0$.

Proof of Theorem 7. We start by proving the first statement. If $r > 1$, then the domain of the CVaR envelope ρ_β is a reflexive Banach space. This, along with the qualification condition given in Assumption 2 implies that the infimal convolution is exact (see [49] and [45, Theorem 3.4]).

[3] A precise definition of Mosco convergence can be found in [48].

The second and third statements follow by direct application of [45, Theorems 2.3 and 2.4]. For the fourth statement we observe that infimal convolution preserves convexity and it holds that $\rho_\beta^* = \rho^* + (\mathrm{CVaR}^\beta)^*$ (e.g. [45, Theorems 3.1, 3.2]). Utilizing the uncertainty set of CVaR, we also know that $(\mathrm{CVaR}^\beta)^* = \delta_{\mathbb{A}_\beta}$ (i.e. it is an indicator function). The proof of this statement then follows by employing [45, Theorem 3.3].

It remains to show the fifth statement. Firstly, we show that ρ_β is coherent if ρ is positively homogeneous. As already discussed, convexity is immediately satisfied. The rest of the proof follows by direct application of [45, Theorems 2.4, 2.5, 3.2 and 3.3]. In particular, we observe that CVaR^β is a continuous and coherent risk measure, while $\mathrm{dom}\ \rho + \mathrm{dom}\ \mathrm{CVaR}^\beta$ is in fact $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$. Moreover, since ρ is positively homogeneous (and using Assumption 2), we obtain that

$$\rho_\beta(Z) = \max_{\zeta \in \widehat{\mathbb{A}}_\beta} \langle \zeta, Z \rangle ,$$

where

$$\widehat{\mathbb{A}}_\beta = \left\{ \zeta \in \mathbb{A} \middle| \zeta \leq 1/\beta,\ \zeta \geq 0,\ \mathbb{E}\{\zeta\} = 1 \right\},$$

and $\mathbb{A}$ is the uncertainty set of ρ. Upon noting that the set $\widehat{\mathbb{A}}_\beta$ is weakly*-closed (e.g., see [1, Chapter 6]), we obtain that ρ_β (as a support function) is weakly lower semicontinuous (see [46, Proposition 3.2.3]). Monotonicty and translation equivariance can be trivially shown (see [50, Section 3.3]). By convexity we obtain that $\left(\rho^* + \mathrm{CVaR}_\beta^*\right)^* = \mathrm{lsc}\,(\rho_\beta)$ (where $\mathrm{lsc} f$ denotes the lower semicontinuous hull of f, in this case with respect to the weak topology), and $\mathrm{lsc}\,(\rho_\beta) = \rho_\beta$ follows by weak lower semicontinuity. To complete the proof, we observe that if ρ is coherent, then $\widehat{\mathbb{A}}_\beta = \left\{ \zeta \in \mathbb{A} \middle| \zeta \leq 1/\beta \right\}$ and the sequence of increasing sets $\{\widehat{\mathbb{A}}_\beta\}_{\beta \searrow 0}$ converges in the Kuratowski-Painlevé sense (see [42, Chapter 4.B]), with respect to both the strong and the weak topology of $\mathcal{L}_r(\mathrm{P}, \mathbb{R})$, to $\mathrm{cl}(\mathbb{A})$ (where the closure is taken with respect to the strong and weak topology in each case). Upon noting that $\mathrm{cl}(\mathbb{A}) = \mathbb{A}$ (by weak*-closedness), we observe that $\delta_{\widehat{\mathbb{A}}_\beta}$ Mosco-converges to $\delta_{\mathbb{A}}$ as $\beta \searrow 0$ (indeed, if a sequence of sets converges to another set in the Kuratowski-Painlevé sense with respect to both the strong and weak sense, then it also converges in the Mosco sense, see [38]). Moreover, since $\rho = \delta_{\mathbb{A}}^*$ and $\rho_\beta = \delta_{\widehat{\mathbb{A}}_\beta}^*$, we obtain that ρ_β^* Mosco-converges, in the functional sense (see [51]), to ρ^* on $\mathcal{L}_r(\mathrm{P}, \mathbb{R})$. But since $r = p$ and $p > 1$, the underlying Banach space is reflexive, and thus using the fact that $\rho_\beta^* \to_M \rho^*$ if and only if $\rho_\beta \to_M \rho$ (the proof of which fact can be found in [48]), we obtain that ρ_β converges, in the Mosco sense, to ρ. ■

Let us now consider an instance of (RCP) for which the associated risk measures are coherent and real-valued with domains that are strict subsets of $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$, say $\mathcal{L}_p(\mathrm{P}, \mathbb{R})$ with some $p > 1$. In this case, it is reasonable to assume that $\boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot) \in \mathcal{L}_p(\mathrm{P}, \mathbf{C})$. We note that Theorem 1 is not directly applicable to this setting. Still, we could instead pose a closely related approximation, involving the CVaR envelopes of the associated risk measures, ensuring that, on the one hand, the approximating problem is now guaranteed to exhibit strong duality, while, on the other hand, the approximation accuracy is controlled by the CVaR level of each corresponding CVaR envelope. Indeed, by utilizing the equivalent reformulation of (RCP) given in (RCP-E), and by letting the associated risk measures admit a dual representation with uncertainty sets $\mathbb{A}^S \triangleq \mathbb{A}^1 \times \ldots \times \mathbb{A}^N$, where $\mathbb{A}^i \subseteq \mathcal{L}_q(\mathrm{P}, \mathbb{R}), i \in \mathbb{N}_N^+, 1/p + 1/q = 1$, such an approximation (at level $\beta \in (0, 1]$) reads as

$$\begin{array}{ll} \underset{\boldsymbol{x}, \boldsymbol{p}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq \inf_{\boldsymbol{\zeta} \in \mathbb{A}_\beta^S} \mathbb{E}\{\boldsymbol{\zeta}(\boldsymbol{H}) \circ \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})\} \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0} \\ & (\boldsymbol{x}, \boldsymbol{p}) \in \mathcal{X} \times \Pi \end{array} \Leftrightarrow \begin{array}{ll} \underset{\boldsymbol{x}, \boldsymbol{p}(\cdot), \boldsymbol{y}(\cdot)}{\text{maximize}} & g^o(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{x} \leq -\boldsymbol{\rho}\left(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) - \boldsymbol{y}(\boldsymbol{H})\right) \\ & \qquad - \mathrm{CVaR}^\beta(\boldsymbol{y}(\boldsymbol{H})) \\ & \boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0} \\ & (\boldsymbol{x}, \boldsymbol{p}, \boldsymbol{y}) \in \mathcal{X} \times \Pi \times \mathcal{L}_1(\mathrm{P}, \mathbf{C}) \end{array},$$

where $\mathbb{A}_\beta^S \triangleq \mathbb{A}_\beta^1 \times \ldots \times \mathbb{A}_\beta^N$, with $\mathbb{A}_\beta^i \triangleq \{\zeta \in \mathbb{A}^i \big| |\zeta(\cdot)| \leq 1/\beta\}, i \in \mathbb{N}_N^+$. Note that we used Theorem 7 (implicitly utilizing Assumption 2) to construct the new uncertainty sets of the associated CVaRized risk measures, while, without loss of generality, we used the same CVaR level for each risk measure. Under this framework, and by utilizing Mosco convergence of our approximating sequence of problems, we can study the consistency of this approximating sequence, obtaining conditions under which the minimizers of the approximating problem-sequence eventually converge to some minimizer of the original problem, as $\beta \searrow 0$. While this is left for future work, the reader is referred to the discussions in [50, Section 4] and [42, Chapter 7.E], where similar results have been considered.

6.3 Efficient Frontiers of Mean-Risk Models

Mean-risk models, see, e.g., [1, Section 6.2] and [39, Section 2], are inevitably related to the risk-averse functional program

$$\inf_{Z\in\mathbb{Z}} \{\rho(Z;c) \triangleq \mathbb{E}\{Z\} + c\mathbb{D}\{Z\}\}, \tag{MR}$$

where $Z : \mathcal{H} \to \mathbb{R}$ denotes the position of a decision maker, $\mathbb{D} : \mathcal{L}_1(\mathrm{P}, \mathbb{R}) \to \mathbb{R}$ serves as a functional measuring *statistical dispersion*, and where we tacitly take the feasible set of positions $\mathbb{Z}$ as some decomposable subset of $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$. In [1], the mean-risk approach is justified as the scalarization of a problem where two objectives, namely the mean –$\mathbb{E}\{Z\}$– and the dispersion –$\mathbb{D}\{Z\}$– need to be efficiently balanced by properly choosing an "optimal" feasible position Z. By leveraging our results, we can provide an alternative –and perhaps more rigorous– interpretation of mean-risk models and the associated risk-averse problem (MR) through the lens of Lagrangian duality.

Next, we assume that $\mathbb{D}$ is a (real-valued) convex, lower semicontinuous (thus continuous) and positively homogeneous functional on $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$, and that P is a nonatomic Borel measure. Note also the generality of problem (MR) relative to the nature of the feasible set $\mathbb{Z}$: If $Z(\cdot) \equiv Z_{\boldsymbol{p}}(\cdot) \triangleq f(\boldsymbol{p}(\cdot), \cdot)$ for an arbitrary but integrable f when $\boldsymbol{p}$ is in some decomposable space Π, then $\mathbb{Z}$ can be taken as the *range of* $f(\boldsymbol{p}(\cdot), \cdot)$ *on* Π, which is necessarily decomposable. Indeed, for each pair $Z \in \mathbb{Z}$ and $Z' \in \mathbb{Z}$ and any Borel set $E \in \mathscr{B}(\mathcal{H})$, it is plain that

$$\begin{aligned} Z(\boldsymbol{h})\mathbb{1}_E(\boldsymbol{h}) + Z'(\boldsymbol{h})\mathbb{1}_{\mathcal{H}\setminus E}(\boldsymbol{h}) &= f(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\mathbb{1}_E(\boldsymbol{h}) + f(\boldsymbol{p}'(\boldsymbol{h}), \boldsymbol{h})\mathbb{1}_{\mathcal{H}\setminus E}(\boldsymbol{h}) \\ &= f(\boldsymbol{p}^o(\boldsymbol{h}), \boldsymbol{h}), \end{aligned}$$

where $\boldsymbol{p}^o(\boldsymbol{h}) = \boldsymbol{p}(\boldsymbol{h})\mathbb{1}_E(\boldsymbol{h}) + \boldsymbol{p}'(\boldsymbol{h})\mathbb{1}_{\mathcal{H}\setminus E}(\boldsymbol{h}) \in \Pi$. Therefore, $\mathbb{Z}$ is decomposable.

It then follows that, under such general conditions, the risk-averse problem (MR) is in a well-defined sense equivalent to the risk-constrained problem

$$\begin{aligned} \mathsf{P}^* \triangleq \ &\text{minimize } \ \mathbb{E}\{Z\} \\ &\text{subject to } \ \mathbb{D}\{Z\} \le \varepsilon \\ &\qquad\qquad\quad Z \in \mathbb{Z} \end{aligned} \tag{MR$'$}$$

for some fixed $\varepsilon \in \mathbb{R}$, provided that the dispersion constraint is strictly feasible, i.e., Slater's condition is satisfied. In this version of the decision-making task, the decision-maker tries to minimize their expected costs or losses, while explicitly keeping the corresponding dispersion under control, i.e., under some number ε. By strong duality, it follows that, as long as it is finite, the value of problem (MR$'$) coincides, up to a constant translation (controlled by ε), with that of the *optimal Lagrangian relaxation*

$$\inf_{Z\in\mathbb{Z}} \mathbb{E}\{Z\} + c^*\mathbb{D}\{Z\},$$

where number $c^* \equiv c^*(\varepsilon) \ge 0$ is an optimal dual variable corresponding to (MR$'$) –representing the *optimal price of risk* [1, Section 6.2.1]–, for some specific choice of ε. In other words, it is true that

$$\sup_{c\ge 0} \inf_{Z\in\mathbb{Z}} \rho(Z;c) - c\varepsilon = \sup_{c\ge 0} \Big\{ -c\varepsilon + \inf_{Z\in\mathbb{Z}} \rho(Z;c) \Big\} = -c^*(\varepsilon)\varepsilon + \inf_{Z\in\mathbb{Z}} \rho(Z;c^*(\varepsilon)) = \mathsf{P}^*,$$

and of course $\rho(Z;c^*(\varepsilon)) = \mathbb{E}\{Z\} + c^*(\varepsilon)\mathbb{D}\{Z\}$. Consequently, when we postulate problem (MR), i.e., $\inf_{Z\in\mathbb{Z}} \rho(Z;c^*(\varepsilon))$, we know that its set of optimal solutions contains at least one which *guarantees* a dispersion of level at most ε (provided that P^* is attained, see, e.g., [2, Theorem 4]). If this set happens to be a singleton (possibly up to a measure-zero equivalence class), then the dispersion constraint is satisfied without extra effort.

Well-known risk measures, such as the CVaR, the MAD, spectral risk measures and more generally all law-invariant coherent risk measures on $\mathcal{L}_1(\mathrm{P}, \mathbb{R})$ through their Kusuoka representations [1, Section 6.3.5], and CVaRizations of such risk measures on $\mathcal{L}_p(\mathrm{P}, \mathbb{R})$, for $p \in [1, \infty)$, as discussed earlier in Section 6.2, admit this plausible interpretation as mean-risk models by their construction. This follows as a result of strong duality of (MR′), which is ensured by Theorem 1.

7 Applications

The base problem (RCP) is general, subsuming as special cases various useful problems arising in several applications. In this section, we discuss a small subset of such applications in more detail to highlight this fact. More specifically, we demonstrate how (risk-constrained versions of) certain problems arising in *wireless systems resource allocation*, and *nonconvex risk-constrained learning* can be cast into the form of (RCP). It follows that, as long as Assumption 1 is in effect, all such problems exhibit strong duality, as a consequence of Theorem 1.

7.1 Risk-Constrained Resource Allocation in Wireless Communications

As discussed in Section 1.1, problem (RCP) is canonically motivated by resource allocation problems from wireless systems engineering. In this setting, the random vector $\boldsymbol{H} \in \mathcal{H} \subseteq \mathbb{R}^{N_{\boldsymbol{H}}}$ quantifies the quality of links between communicating or networking entities (nodes), $\boldsymbol{p}$ represents a resource allocation policy, and $\boldsymbol{f}$ measures instantaneous service levels achieved by a policy $\boldsymbol{p}$ operating on the observable $\mathcal{H}$. The risk of $\boldsymbol{f}$ is associated with the *ergodic performance* vector $\boldsymbol{x} \in \mathbb{R}^N$, and utilities g° and $\boldsymbol{g}$ are used to evaluate those ergodic risks. In fact, there are numerous resource allocation tasks arising in practical settings that are readily modelled via (RCP), primarily considered in the risk-neutral setting (see, e.g., [2, 10, 12–15, 19, 21, 32]). Below we consider two related examples to showcase the expressive power of (RCP) in the risk-constrained setting.

7.1.1 Multiple Access Interference Channel

We first develop a generalization of the model studied in, e.g., [19, Section VI.B], [21, Section VI] and [52], where a *multiple access interference channel* model is considered, in which there are $N_S \equiv N_{\boldsymbol{H}}$ wireless transmitters simultaneously communicating with a central node, e.g., a base station. Here, each component of the uncertain element $\boldsymbol{H}$ corresponds to the strength of the communication channel between each transmitter and the base station. The signal emitted by each transmitter introduces interference to all other signals emitted by the remaining transmitters, and the goal is to optimally allocate power to each transmitter on the basis of observing the channel information vector $\boldsymbol{H}$, so as to maximize a certain Quality-of-Service (QoS) network-wide utility, under a total expected power specification $p_{max} > 0$.

More specifically, consider the hollow matrix $\boldsymbol{T} \triangleq \mathbf{1}\mathbf{1}^\top - \boldsymbol{I}$ (where $\mathbf{1}$ denotes the vector of ones of appropriate dimension), and define the service vector stacking the communication rates achievable by all transmitters at the base station, i.e.,

$$\boldsymbol{f}_S(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) \triangleq \log\left(\mathbf{1} + \frac{\boldsymbol{H} \circ \boldsymbol{p}(\boldsymbol{H})}{\sigma^2 \boldsymbol{I} + \boldsymbol{T}[\boldsymbol{H} \circ \boldsymbol{p}(\boldsymbol{H})]}\right), \quad \boldsymbol{p} \in \Pi,$$

where the $\log(\cdot)$ and division operations on vectors are interpreted component-wise, $\sigma^2 > 0$ denotes the variance of the (common) reception noise induced by the transmission to the base station, and Π is any decomposable space of nonnegative resource policies; note that $\boldsymbol{p}$ represents power. Observe that $\boldsymbol{f}_S(\cdot, \boldsymbol{H})$ is component-wise nonconcave and nonlinearly coupled, i.e., the communication rate of each transmitter depends on the power allocations of all transmitters in the network. We also define the scalar resource coupling function $f_C(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}) \triangleq p_{max} - \langle \boldsymbol{p}(\boldsymbol{H}), \mathbf{1}\rangle, \boldsymbol{p} \in \Pi$. By choosing for simplicity $g^\circ(\boldsymbol{x}) = \langle \boldsymbol{\alpha}, \boldsymbol{x}\rangle$, i.e., a linear *sumrate utility* with nonnegative weights $\boldsymbol{\alpha} \geq \mathbf{0}$ evaluating the *(risk-)ergodic* service vector $\boldsymbol{x} \in \mathbb{R}^{N_S}$, the corresponding stochastic resource allocation problem studied in [19, 21, 52] admits a risk-constrained generalization which reads as

$$\begin{aligned}
\underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{maximize}} \quad & \langle \boldsymbol{\alpha}, \boldsymbol{x} \rangle \\
\text{subject to} \quad & \boldsymbol{x} \leq -\boldsymbol{\rho}_S \left[-\log \left(\mathbf{1} + \frac{\boldsymbol{H} \circ \boldsymbol{p}(\boldsymbol{H})}{\boldsymbol{\nu} + \boldsymbol{T}[\boldsymbol{H} \circ \boldsymbol{p}(\boldsymbol{H})]} \right) \right], \\
& \langle \mathbb{E}\{\boldsymbol{p}(\boldsymbol{H})\}, \mathbf{1} \rangle \leq p_{max} \\
& (\boldsymbol{x}, \boldsymbol{p}) \in \mathbb{R}_+^{N_S} \times \Pi
\end{aligned} \tag{MAI}$$

where $\boldsymbol{\rho}_S$ is a finite-valued vector risk measure on $\mathcal{L}_1(\mathrm{P}, \mathbb{R}^{N_S})$ and compatible with the construction of (RCP); for instance, $\boldsymbol{\rho}_S$ could be a vector of CVaRs with potentially distinct confidence levels.

Evidently, problem (MAI) is a special case of (RCP), since the former can be rewritten as

$$\begin{aligned}
\underset{(\boldsymbol{x}_S, x_C), \boldsymbol{p}(\cdot)}{\text{maximize}} \quad & g^o(\boldsymbol{x}_S) \\
\text{subject to} \quad & \begin{bmatrix} \boldsymbol{x}_S \\ x_C \end{bmatrix} \leq \begin{bmatrix} -\boldsymbol{\rho}_S(-\boldsymbol{f}_S(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})) \\ -\mathbb{E}\{-f_C(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}))\} \end{bmatrix} \triangleq -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H})), \\
& (\boldsymbol{x}_S, x_C, \boldsymbol{p}) \in \mathcal{X} \times \Pi
\end{aligned}$$

with the identifications $\boldsymbol{f} = (\boldsymbol{f}_S, f_C)$ and $\mathcal{X} = \mathbb{R}_+^{N_S} \times \{0\} \subset \mathbb{R}^{N = N_S + 1}$. Consequently, provided integrability of $\boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot)$, (MAI) exhibits strong duality under mere nonatomicity of the underlying Borel measure P associated with the channel vector $\boldsymbol{H}$, in combination with Slater's constraint qualification; those conditions suffice to ensure that Assumption 1 holds. Let us note that despite using a simple linear utility g°, (MAI) remains a highly nontrivial and challenging nonconvex problem.

7.1.2 Frequency Division Broadcast Channel

We next consider the optimal risk-aware design of a *frequency division broadcast channel* model, generalizing another risk-neutral problem studied in [2]. In this case, an Access Point (AP) is tasked to optimally allocate limited resources in order to communicate with $N_S = N_{\boldsymbol{H}}$ terminals, or users; this sometimes is called a *broadcast* or *downlink* setting, since the AP transmits information to all terminals. As before, each of the components of $\boldsymbol{H}$ corresponds to the channel strength between the AP and each terminal. However, in this model the AP exploits *frequency division multiplexing*, which allows simultaneous transmission to multiple terminals over different frequencies (carriers, or bands), therefore with no cross-interference. The goal of the AP is to *jointly* select frequencies over which to transmit *and* allocate power to each set of selected transmissions (one for each frequency selected), with *both* decisions made on the basis of observing the channel vector $\boldsymbol{H}$, so as to maximize network-wide QoS, under a total expected power specification $p_{max} > 0$.

Power and frequency allocation policies are denoted by $\boldsymbol{p} \in \Pi$ and $\boldsymbol{\phi} \in \Phi_{N_A}$ for some $N_A \leq N_S$, respectively, where Π any decomposable set of nonnegative policies, and Φ_{N_A} is defined as

$$\Phi_{N_A} = \left\{ \boldsymbol{\phi} \in \mathcal{L}_\infty(\mathrm{P}, \mathbb{R}^{N_S}) \middle| \boldsymbol{\phi}(\cdot) \in \{0,1\}^{N_S},\ \langle \boldsymbol{\phi}(\cdot), \mathbf{1} \rangle \leq N_A,\ \text{P-a.e.} \right\}.$$

In other words, each element $\boldsymbol{\phi} \in \Phi_{N_A}$ indicates which transmissions are active –these are *at most* N_A out of N_S–, each of them realized over the corresponding frequency. It is easy to verify that for any choice of the number of active transmissions N_A, the set Φ_{N_A} is decomposable. By assuming practical Adaptive Modulation and Conding (AMC) with $M \in \mathbb{N}^+$ modes, the service vector stacking the communication rates achievable at each terminal takes the form [2, Section on "Frequency Division Broadcast Channel"]

$$\boldsymbol{f}_S([\boldsymbol{\phi}(\boldsymbol{H}), \boldsymbol{p}(\boldsymbol{H})], \boldsymbol{H}) \triangleq \sum_{m=1}^{M} q_m \left[\boldsymbol{\phi}(\boldsymbol{H}) \circ \mathbb{1}_{[l_m, l_{m+1})} \left(\frac{\boldsymbol{H} \circ \boldsymbol{p}(\boldsymbol{H})}{\boldsymbol{\nu}} \right) \right], \quad (\boldsymbol{\phi}, \boldsymbol{p}) \in \Phi_{N_A} \times \Pi,$$

where $\boldsymbol{\nu} > \mathbf{0}$ contains the variances of the reception noises (possibly) induced at each of the terminals, and where $q_m \geq 0$ is the m-th mode rate supported at a terminal, which is achieved whenever the corresponding received signal-to-noise ratio –i.e., the fractional term inside the indicator function in the expression above– is between predefined levels $l_m \geq 0$ and $l_{m+1} \geq 0$ (with $l_m < l_{m+1}$, of course), for $m \in \mathbb{N}_M^+$; there are M possible operation modes of this form in total. Similarly to Section 7.1.1, we also define the coupling function $f_C([\boldsymbol{\phi}(\boldsymbol{H}), \boldsymbol{p}(\boldsymbol{H})], \boldsymbol{H}) \triangleq p_{max} - \langle \boldsymbol{\phi}(\boldsymbol{H})\boldsymbol{p}(\boldsymbol{H}), \mathbf{1}\rangle$, $(\boldsymbol{\phi}, \boldsymbol{p}) \in \Phi_{N_A} \times \Pi$.

Lastly, we consider any strictly increasing concave utility u evaluating each component of the (risk-)ergodic service vector $\boldsymbol{x} \in \mathbb{R}^{N_S}$ –we succintly write $\langle u(\boldsymbol{x}), \mathbf{1}\rangle \triangleq g^o(\boldsymbol{x})$–, the latter are further constrained within the box compactum $\mathcal{B} \triangleq \left\{\boldsymbol{x} \in \mathbb{R}^{N_S} \middle| x_{min}\mathbf{1} \leq \boldsymbol{x} \leq x_{max}\mathbf{1}\right\}$, for fixed numbers $0 \leq x_{min} < x_{max}$. Then, the corresponding stochastic resource allocation problem may be formulated as

$$\begin{aligned} \underset{\boldsymbol{x}, \boldsymbol{\phi}(\cdot), \boldsymbol{p}(\cdot)}{\text{maximize}} \quad & \langle u(\boldsymbol{x}), \mathbf{1}\rangle \\ \text{subject to} \quad & \boldsymbol{x} \leq -\boldsymbol{\rho}_S\left[-\sum_{m=1}^{M} \pi_m \left[\boldsymbol{\phi}(\boldsymbol{H}) \circ \mathbb{1}_{[l_m, l_{m+1})}\left(\frac{\boldsymbol{H} \circ \boldsymbol{p}(\boldsymbol{H})}{\boldsymbol{\nu}}\right)\right]\right], \\ & \langle \mathbb{E}\{\boldsymbol{\phi}(\boldsymbol{H})\boldsymbol{p}(\boldsymbol{H})\}, \mathbf{1}\rangle \leq p_{max} \\ & (\boldsymbol{x}, \boldsymbol{\phi}, \boldsymbol{p}) \in \mathcal{B} \times \Phi_{N_A} \times \Pi \end{aligned} \tag{FDB}$$

where $\boldsymbol{\rho}_S$ is a finite-valued vector risk measure on $\mathcal{L}_1(\mathrm{P}, \mathbb{R}^{N_S})$ and compatible with the construction of (RCP), as in Section 7.1.1. We observe that (FDB) is a highly nonconvex functional mixed-integer program; in fact, the policy $\boldsymbol{\phi}$ is integer-valued, and the service function evaluated by the risk measure $\boldsymbol{\rho}_S$ is clearly discontinuous.

Again, problem (FDB) is a special case of (RCP), because we can restate the former as

$$\begin{aligned} \underset{(\boldsymbol{x}_S, x_C), \boldsymbol{p}'(\cdot)}{\text{maximize}} \quad & g^o(\boldsymbol{x}_S) \\ \text{subject to} \quad & \begin{bmatrix} \boldsymbol{x}_S \\ x_C \end{bmatrix} \leq \begin{bmatrix} -\boldsymbol{\rho}_S(-\boldsymbol{f}_S(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})) \\ -\mathbb{E}\{-f_C(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})\} \end{bmatrix} \triangleq -\boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H})), \\ & (\boldsymbol{x}_S, x_C, \boldsymbol{p}') \in \mathcal{X} \times \Pi' \end{aligned}$$

with the identifications $\boldsymbol{p}' = (\boldsymbol{\phi}, \boldsymbol{p})$, $\boldsymbol{f} = (\boldsymbol{f}_S, f_C)$, $\Pi' = \Phi_{N_A} \times \Pi$ and $\mathcal{X} = \mathcal{B} \times \{0\} \subset \mathbb{R}^{N = N_S + 1}$, and where we can easily verify that the product Π' is a decomposable set. If $\boldsymbol{f}(\boldsymbol{p}'(\cdot), \cdot)$ is integrable on Π' and P is nonatomic, it follows that problem (FDB) exhibits strong duality as long as it is strictly feasible. This is true despite the discontinuity of $\boldsymbol{f}(\cdot, \boldsymbol{H})$ and the nonconvexity of the set of feasible frequency policies policies Φ_{N_A}.

7.2 Risk-Constrained Learning with Nonconvex Losses

Next, we consider a general formulation in the context of supervised risk-constrained learning, where the associated loss functions are allowed to be nonconvex. In the risk-neutral setting, this class of problems has been recently studied in [23]. On our usual probability space $(\Omega, \mathscr{F}, \mu)$, we consider random *example vectors* $(\boldsymbol{X}_i, Y_i) \colon \Omega \to \mathbb{R}^d \times \mathbb{R}$, $i \in \mathbb{N}_m$ together with their induced Borel probability distributions $\mathcal{D}_i : \mathscr{B}(\mathbb{R}^d \times \mathbb{R}) \to [0, 1]$, instantiated over data pairs $(\boldsymbol{x}, y)$, where $\boldsymbol{x} \in \mathbb{R}^d$ represents a realized feature or system input and $y \in \mathbb{R}$ represents a realized label or measurement. We denote by $\mathcal{D}_{\boldsymbol{X}_i}$ the marginal distribution of feature $\boldsymbol{X}_i$, and by $\mathcal{D}_{Y_i|\boldsymbol{X}_i}$ the conditional distribution of label Y_i given feature $\boldsymbol{X}_i$, for $i \in \mathbb{N}_m$.

We assume that the marginal probability distributions $\mathcal{D}_{\boldsymbol{X}_i}$, for $i \in \mathbb{N}_m^+$, are absolutely continuous with respect to a "common denominator" $\mathcal{D}_{\boldsymbol{X}_0}$ (without loss of generality), which in turn is assumed to be nonatomic. Consequently, the Radon-Nikodym theorem implies existence of integrable functions $w_i \colon \mathbb{R}^d \to \mathbb{R}^+$, such that $w_i \triangleq \mathrm{d}\mathcal{D}_{\boldsymbol{X}_i} / \mathrm{d}\mathcal{D}_{\boldsymbol{X}_0}$, for all $i \in \mathbb{N}_m^+$.

Letting $\ell_i \colon \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}_+$ be given (possibly nonconvex) loss functions, for $i \in \mathbb{N}_m$, we consider the risk-over-risk functional constrained learning problem

$$\begin{array}{ll} \underset{\boldsymbol{f}(\cdot)\in\mathcal{F}}{\text{minimize}} & \rho_0\big(\widehat{\rho}_0\big(\ell_0\big(\boldsymbol{f}(\boldsymbol{X}_0), Y_0\big)\big|\boldsymbol{X}_0\big)\big) \\ \text{subject to} & \rho_i\big(\widehat{\rho}_i\big(\ell_i\big(\boldsymbol{f}(\boldsymbol{X}_i), Y_i\big)\big|\boldsymbol{X}_i\big)\big) \le c_i, \quad i \in \mathbb{N}_m^+ \end{array}, \tag{RCL}$$

where, for $i \in \mathbb{N}_m$, ρ_i is a convex, lower semicontinuous, and positively homogeneous risk measure taking finite values on $\mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$, while $\widehat{\rho}_i(\cdot|\boldsymbol{X}_i)$ is a $\mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$-valued conditional risk measure over the conditional measure of Y_i given $\boldsymbol{X}_i$, which is merely assumed to obey a *substitution rule* of the form

$$\widehat{\rho}_i(\ell_i(\boldsymbol{f}(\boldsymbol{X}_i), Y_i)|\boldsymbol{X}_i) \equiv \widehat{\rho}_i(\ell_i(\boldsymbol{z}, Y_i)|\boldsymbol{X}_i)\big|_{\boldsymbol{z}=\boldsymbol{f}(\boldsymbol{X}_i)} \equiv F_i\left(\boldsymbol{f}(\boldsymbol{X}_i), \boldsymbol{X}_i\right),$$

where $F_i(\boldsymbol{f}(\cdot), \cdot) \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$ for any *learning representation* (i.e., policy) $\boldsymbol{f} : \mathbb{R}^d \to \mathbb{R}$, the latter lying in an appropriate decomposable space $\mathcal{F}$. Lastly, $c_i \in \mathbb{R}$, for all $i \in \mathbb{N}_m^+$.

Since $\mathcal{D}_{\boldsymbol{X}_i}$ is absolutely continuous relative to $\mathcal{D}_{\boldsymbol{X}_0}$ and $F_i(\boldsymbol{f}(\cdot), \cdot) \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$ by assumption, it also follows that $G_i(\boldsymbol{f}(\cdot), \cdot) \triangleq F_i(\boldsymbol{f}(\cdot), \cdot)w_i \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_0}, \mathbb{R})$, for all $i \in \mathbb{N}_m^+$. Let $\mathbb{A}^i \subseteq \mathcal{L}_\infty(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$ be the uncertainty set corresponding to ρ_i, $i \in \mathbb{N}_m$. Then, for each $i \in \mathbb{N}_m^+$ we may construct another (related) convex, lower semicontinuous, and positively homogeneous risk measure $\widetilde{\rho}_i$ on $\mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_0}, \mathbb{R})$, which is such that $\widetilde{\rho}_i(Zw_i) = \rho_i(Z)$, for any $Z \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$. In particular, for $i \in \mathbb{N}_m^+$ we define

$$\begin{aligned} \widetilde{\rho}_i\left(Z(\boldsymbol{X}_0)w_i(\boldsymbol{X}_0)\right) &\triangleq \sup_{\zeta\in\mathbb{A}^i} \int \zeta(\boldsymbol{x})Z(\boldsymbol{x})w_i(\boldsymbol{x})\mathrm{d}\mathcal{D}_{\boldsymbol{X}_0}(\boldsymbol{x}) \\ &= \sup_{\zeta\in\mathbb{A}^i} \int \zeta(\boldsymbol{x})Z(\boldsymbol{x})\mathrm{d}\mathcal{D}_{\boldsymbol{X}_i}(\boldsymbol{x}) = \rho_i\left(Z(\boldsymbol{X}_i)\right), \quad Z \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R}), \end{aligned}$$

where we have exploited the dual representation of ρ_i (also note that $\zeta Z w_i \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_0}, \mathbb{R})$). Further, it can be shown that the supremum in the second integral above would not change by restricting $\mathbb{A}^i$ to an appropriately selected bounded set $\widetilde{\mathbb{A}}^i \subseteq \mathcal{L}_\infty(\mathcal{D}_{\boldsymbol{X}_0}, \mathbb{R})$, chosen independently of Zw_i and with $\mathbb{A}^i \supseteq \widetilde{\mathbb{A}}^i$, by carefully noticing that the integral is taken with respect to $\mathcal{D}_{\boldsymbol{X}_i}$, and does not change for functions differing on $\mathcal{D}_{\boldsymbol{X}_i}$-null sets. In other words, $\widetilde{\rho}_i$ admits the desired representation

$$\widetilde{\rho}_i\left(Z(\boldsymbol{X}_0)w_i(\boldsymbol{X}_0)\right) = \sup_{\zeta\in\widetilde{\mathbb{A}}^i} \int \zeta(\boldsymbol{x})[Z(\boldsymbol{x})w_i(\boldsymbol{x})]\mathrm{d}\mathcal{D}_{\boldsymbol{X}_0}(\boldsymbol{x}), \quad Z \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R}),$$

for some bounded risk envelope $\widetilde{\mathbb{A}}^i \subseteq \mathcal{L}_\infty(\mathcal{D}_{\boldsymbol{X}_0}, \mathbb{R})$, also verifying that $\widetilde{\rho}_i$ is a convex, lower semicontinuous and positively homogeneous risk measure on the subset of $\mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_0}, \mathbb{R})$ generated by functions of the form Zw_i, for $Z \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$, $i \in \mathbb{N}_m^+$. Although not necessary, we can also extend $\widetilde{\rho}_i$ everywhere on $\mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_0}, \mathbb{R})$, for each $i \in \mathbb{N}_m^+$. Therefore, it follows that, for every $\boldsymbol{f} \in \mathcal{F}$,

$$\rho_i\big(\widehat{\rho}_i\big(\ell_i\big(\boldsymbol{f}(\boldsymbol{X}_i), Y_i\big)\big|\boldsymbol{X}_i\big)\big) = \rho_i\left(F_i\big(\boldsymbol{f}(\boldsymbol{X}_i), \boldsymbol{X}_i\big)\right) = \widetilde{\rho}_i\left(G_i\big(\boldsymbol{f}(\boldsymbol{X}_0), \boldsymbol{X}_0\big)\right),$$

and we can then equivalently recast (RCL) as (under our assumptions)

$$\begin{array}{ll} \underset{\boldsymbol{f}(\cdot)\in\mathcal{F}}{\text{minimize}} & \rho_0\left(F_0\big(\boldsymbol{f}(\boldsymbol{X}_0), \boldsymbol{X}_0\big)\right) \\ \text{subject to} & \widetilde{\rho}_i\left(G_i\big(\boldsymbol{f}(\boldsymbol{X}_0), \boldsymbol{X}_0\big)\right) \le c_i, \quad i \in \mathbb{N}_m^+ \end{array}. \tag{RCL$'$}$$

Note that (RCL′) is now an instance of (RCP) –Section 6.1–, and hence Slater's condition suffices to ensure that (RCL′) satisfies Assumption 1. Theorem 1 then implies that (RCL′) exhibits strong duality, and the same of course holds for (RCL).

At this point, we have described problem (RCL) in full generality. For the convenience of the reader we now discuss specific instances of (RCL) which can be equivalently expressed in the form of (RCL′), to demonstrate the compatibility and generality of our assumptions. Firstly, let us notice that our assumption of absolute continuity of the marginal distributions $\mathcal{D}_{\boldsymbol{X}_i}$ with respect to $\mathcal{D}_{\boldsymbol{X}_0}$ or more generally some other common denominator measure is quite standard in the literature. Indeed, it holds if each distribution $\mathcal{D}_{\boldsymbol{X}_i}, i \in \mathbb{N}_m$ is assumed to admit a density relative to the Lebesgue measure (see, e.g., [23]), which is of course nonatomic, considered on a compact subset of $\mathbb{R}^d$. The latter minor technicality just imposes boundedness on the features, and is essential for keeping the Lebesgue measure finite, thus rendering it equivalent to a probability measure. Additionally, assuming nonatomicity of $\mathcal{D}_{\boldsymbol{X}_0}$ itself or any of the rest of the $\mathcal{D}_{\boldsymbol{X}_i}, i \in \mathbb{N}_m^+$ is standard (see, e.g., [23, Theorem 1]) and intuitive, since the features are for a fact continuously valued in numerous machine learning applications.

Our second assumption concerns the substitution rule for the conditional risk mappings $\widehat{\rho}_i(\cdot|\boldsymbol{X}_i)$. This is not restrictive and can be shown to hold for all convex, lower semicontinuous, and positively homogeneous risk measures of interest herein. In fact, the substitution rule is quite general and holds beyond the aforementioned class. For instance, several compositional risk measures studied in [53] satisfy the substitution rule, being in particular nonhomogeneous.

Finally, let us specialize our result to the case where $\rho_i(\cdot) = \mathbb{E}_{\mathcal{D}_{\boldsymbol{X}_i}}\{\cdot\}$ and $\widehat{\rho}_i(\cdot|\boldsymbol{X}_i) = \mathbb{E}_{\mathcal{D}_{Y_i|\boldsymbol{X}_i}}\{\cdot\}$, for all $i \in \mathbb{N}_m$. This setting corresponds to risk-neutral constrained learning considered in [23]. Assuming that $\mathcal{D}_{\boldsymbol{X}_0}$ is nonatomic and $\mathcal{D}_{\boldsymbol{X}_0} \gg \mathcal{D}_{\boldsymbol{X}_i}$ for all $i \in \mathbb{N}_m^+$ (or any alternative situation; see above), together with Slater's constraint qualification and that $\mathbb{E}\left\{\ell_i\left(\boldsymbol{f}(\boldsymbol{X}_i), Y_i\right)|\boldsymbol{X}_i\right\} \in \mathcal{L}_1(\mathcal{D}_{\boldsymbol{X}_i}, \mathbb{R})$ for every $\boldsymbol{f} \in \mathcal{F}$, where $\mathcal{F}$ is a decomposable space, we recover the results of [23, Proposition III.2 and Proposition B.1], unifying the classification and regression regimes, while dispensing certain assumptions. In particular, the zero duality gap result on regression problems given in [23, Proposition B.1] is shown here to hold without the requirement postulated in [23, Assumption 6].

8 Conclusions

We established strong duality for a wide class of risk-constrained nonconvex functional programs. Our technical approach has exploited Uhl's extension of Lyapunov's convexity theorem for Banach-valued vector-measures together with risk duality, and is applicable to programs involving convex, lower semicontinuous and positively homogeneous risk measures on $\mathcal{L}_1$, strictly generalizing existing results available for the risk-neutral setting, without imposing additional assumptions. We further discussed extensions of our results covering a wider class of problems, implications on the interpretation of mean-risk models, as well as certain limitations of our results, possibly identifying topics for further investigation. Lastly, we showcased the applicability of the theory on two specific application areas with practical interest, namely, risk-constrained wireless systems resource allocation, and risk-constrained learning with nonconvex losses. In the latter case, we have recovered known strong duality results for risk-neutral constrained learning under relaxed assumptions, while also emphasizing the expressive power of the risk-constrained model studied herein.

Acknowledgments. The authors would like to kindly acknowledge support by a Microsoft gift, as well as by the US National Science Foundation (NSF) under Grant CCF 2242215.

Appendix A Bochner Integral Representation of $\boldsymbol{G_p}$

To verify the seemingly obvious though not immediate equivalence

$$\boldsymbol{G_p}(E) = \int_E \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}), \quad E \in \mathscr{B}(\mathcal{H}),$$

we first need to show that the function $\boldsymbol{\Lambda}_{\mathbb{B}}(\cdot) \otimes \boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot) \in \mathbb{X} = \ell_\infty$ is strongly Bochner-measurable. By definition [29, Definition II.1.1], we need to verify the existence of a sequence of simple $\mathbb{X}$-valued functions $\{\boldsymbol{g}^m : \mathcal{H} \to \mathbb{X}\}_{m \in \mathbb{N}}$ such that

$$\lim_{m \to \infty} \|\boldsymbol{g}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\mathbb{X}} = 0, \quad \text{P-a.e.}$$

First, observe that, for every $\boldsymbol{h} \in \mathcal{H}$, each of the elements of $\boldsymbol{\Lambda}_{\mathbb{B}}$ may be written as

$$\lambda_n(\boldsymbol{h}) \triangleq \sum_{i \in \mathcal{I}_n \subset \mathbb{N}} \tilde{\varrho}_n^i \mathbb{1}_{D_i}(\boldsymbol{h}) \triangleq \sum_{i=0}^{\infty} \varrho_n^i \mathbb{1}_{D_i}(\boldsymbol{h}), \quad \text{for some finite index set } \mathcal{I}_n, n \in \mathbb{N},$$

with $\tilde{\varrho}_n^i \in \mathbb{Q} \bigcap [-\gamma, \gamma]$ and where we have further defined

$$\varrho_n^i \triangleq \tilde{\varrho}_n^i \mathbb{1}_{\mathcal{I}_n}(i) = \begin{cases} \tilde{\varrho}_n^i & \text{if } i \in \mathcal{I}_n \\ 0, & \text{if not} \end{cases}, \quad (i, n) \in \mathbb{N} \times \mathbb{N}.$$

Consequently, it is true that

$$\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) = \begin{bmatrix} \lambda_0(\boldsymbol{h}) \\ \lambda_1(\boldsymbol{h}) \\ \vdots \\ \lambda_n(\boldsymbol{h}) \\ \vdots \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{\infty} \varrho_0^i \mathbb{1}_{D_i}(\boldsymbol{h}) \\ \sum_{i=0}^{\infty} \varrho_1^i \mathbb{1}_{D_i}(\boldsymbol{h}) \\ \vdots \\ \sum_{i=0}^{\infty} \varrho_n^i \mathbb{1}_{D_i}(\boldsymbol{h}) \\ \vdots \end{bmatrix} = \sum_{i=0}^{\infty} \begin{bmatrix} \varrho_0^i \\ \varrho_1^i \\ \vdots \\ \varrho_n^i \\ \vdots \end{bmatrix} \mathbb{1}_{D_i}(\boldsymbol{h}) \in \mathbb{X}, \quad \boldsymbol{h} \in \mathcal{H},$$

confirming that $\boldsymbol{\Lambda}_{\mathbb{B}}$ is countably valued; by defining $\boldsymbol{\varrho}^i \triangleq [\tilde{\varrho}_0^i \ \tilde{\varrho}_1^i \ \ldots] \in \mathbb{X}$, $\boldsymbol{\Lambda}_{\mathbb{B}}$ can be represented as

$$\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) = \sum_{i=0}^{\infty} \boldsymbol{\varrho}^i \mathbb{1}_{D_i}(\boldsymbol{h}), \quad \boldsymbol{h} \in \mathcal{H}.$$

Now, clip-off $\boldsymbol{\Lambda}_{\mathbb{B}}$ and define the approximating family of simple $\mathbb{X}$-valued functions on $\mathcal{H}$ with members

$$\boldsymbol{\Lambda}_{\mathbb{B}}^m(\boldsymbol{h}) = \sum_{i=0}^{m} \boldsymbol{\varrho}^i \mathbb{1}_{D_i}(\boldsymbol{h}), \quad \boldsymbol{h} \in \mathcal{H},$$

for each $m \in \mathbb{N}$, and consider another standard sequence of finite-dimensional simple functions $\{[\boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot)]^m\}_{m \in \mathbb{N}}$ converging pointwise to $\boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot)$. Since the product of simple functions is also a simple function, let us choose, for every $m \in \mathbb{N}$,

$$\boldsymbol{g}^m(\boldsymbol{h}) \triangleq \boldsymbol{\Lambda}_{\mathbb{B}}^m(\boldsymbol{h}) \otimes [\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m, \quad \boldsymbol{h} \in \mathcal{H}.$$

Apparently, we would like to show that the sequence $\{\boldsymbol{g}^m(\boldsymbol{h})\}_{m \in \mathbb{N}}$ converges in norm to $\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})$ for P-almost every $\boldsymbol{h}$. To do this, we may decompose as (triangle inequality)

$$\begin{aligned} &\|\boldsymbol{g}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\mathbb{X}} \\ &\leq \|\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes ([\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}))\|_{\mathbb{X}} + \|(\boldsymbol{\Lambda}_{\mathbb{B}}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h})) \otimes [\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m\|_{\mathbb{X}}. \end{aligned}$$

For the first term on the right-hand side, we easily have that, for every $\boldsymbol{h} \in \mathcal{H}$,

$$\|\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes ([\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}))\|_{\mathbb{X}} = \|\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h})\|_{\mathbb{X}} \|[\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\infty}$$

and so

$$\lim_{m \to \infty} \|\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes ([\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m - \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}))\|_{\mathbb{X}} = 0, \quad \text{P-a.e.}$$

For the second term, it is similarly true that, for every $\boldsymbol{h} \in \mathcal{H}$,

$$\|(\boldsymbol{\Lambda}_{\mathbb{B}}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h})) \otimes [\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m\|_{\mathbb{X}} = \|\boldsymbol{\Lambda}_{\mathbb{B}}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h})\|_{\mathbb{X}} \|[\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})]^m\|_{\infty},$$

and thus it suffices to look at the approximation $\|\boldsymbol{\Lambda}_{\mathbb{B}}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h})\|_{\mathbb{X}}$. We have

$$\begin{aligned}\|\boldsymbol{\Lambda}_{\mathbb{B}}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h})\|_{\mathbb{X}} &= \left\| \sum_{i=0}^{m} \boldsymbol{\varrho}^i \mathbb{1}_{D_i}(\boldsymbol{h}) - \sum_{i=0}^{\infty} \boldsymbol{\varrho}^i \mathbb{1}_{D_i}(\boldsymbol{h}) \right\|_{\mathbb{X}} \\ &= \left\| \sum_{i=m+1}^{\infty} \boldsymbol{\varrho}^i \mathbb{1}_{D_i}(\boldsymbol{h}) \right\|_{\mathbb{X}} \\ &= \sum_{i=m+1}^{\infty} \|\boldsymbol{\varrho}^i\|_{\mathbb{X}} \mathbb{1}_{D_i}(\boldsymbol{h}) \to 0, \quad \text{as } m \to \infty.\end{aligned}$$

Overall, we have that

$$\lim_{m\to\infty} \|\boldsymbol{g}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\mathbb{X}} = 0, \quad \text{P-a.e.}$$

Because $\boldsymbol{g}^m$ is simple for each $m \in \mathbb{N}$, this shows that $\boldsymbol{\Lambda}_{\mathbb{B}}(\cdot) \otimes \boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot)$ is (strongly) P-measurable.

In particular, it is now plain to see that $\boldsymbol{\Lambda}_{\mathbb{B}}(\cdot) \otimes \boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot)$ is Bochner integrable –and specifically in $\mathcal{L}_1(\mathrm{P}, \mathbb{X})$–; simply observe that, for every $\boldsymbol{h} \in \mathcal{H}$,

$$\begin{aligned}\|\boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\mathbb{X}} &= \sup_{n\in\mathbb{N}} \|\lambda_n(\boldsymbol{h}) \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\infty} \\ &\leq \gamma \|\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\infty}.\end{aligned}$$

Since $\int \|\boldsymbol{f}(\boldsymbol{p}(\cdot), \cdot)\|_{\infty} \mathrm{dP} < \infty$, the claim follows by [29, Theorem II.2.2]. Proceeding in a general fashion, by construction of the Bochner integral there is a sequence of simple $\mathbb{X}$-valued functions $\{\boldsymbol{g}^m : \mathcal{H} \to \mathbb{X}\}_{m\in\mathbb{N}}$ such that [29, Definition II.2.1]

$$\lim_{m\to\infty} \int \|\boldsymbol{g}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\mathbb{X}} \mathrm{dP}(\boldsymbol{h}) = 0.$$

Then, it is true that, for every $(n^o, i^o) \in \mathbb{N} \times \mathbb{N}_N^+$,

$$\begin{aligned}\int \|\boldsymbol{g}^m(\boldsymbol{h}) - \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})\|_{\mathbb{X}} \mathrm{dP}(\boldsymbol{h}) &= \int \sup_{n\in\mathbb{N}} \max_{i\in\mathbb{N}_N^+} |g_{n,i}^m(\boldsymbol{h}) - \lambda_n(\boldsymbol{h}) f_i(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})| \mathrm{dP}(\boldsymbol{h}) \\ &\geq \int |g_{n^o,i^o}^m(\boldsymbol{h}) - \lambda_{n^o}(\boldsymbol{h}) f_{i^o}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})| \mathrm{dP}(\boldsymbol{h}) \geq 0,\end{aligned}$$

which reveals that each component sequence $\{g_{n^o,i^o}^m\}_{m\in\mathbb{N}}$ is such that

$$\lim_{m\to\infty} \int |g_{n^o,i^o}^m(\boldsymbol{h}) - \lambda_{n^o}(\boldsymbol{h}) f_{i^o}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h})| \mathrm{dP}(\boldsymbol{h}) = 0.$$

This of course further implies that, for every $(n, i) \in \mathbb{N} \times \mathbb{N}_N^+$,

$$\lim_{m\to\infty} \int_E g_{n,i}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) = \int_E \lambda_n(\boldsymbol{h}) f_i(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}), \quad E \in \mathscr{B}(\mathcal{H}).$$

Using properties of $\mathbb{X}$ as a vector space (i.e., addition and scalar multiplication), it follows that for every simple function $\boldsymbol{g} : \mathcal{H} \to \mathbb{X}$, we may write, for some fixed vectors $\{\boldsymbol{z}^k \in \mathbb{X}\}_{k\in\mathbb{N}_K^+}$ and events $\{\mathcal{G}_k \in \mathscr{B}(\mathcal{H})\}_{k\in\mathbb{N}_K^+}$ (and with the obvious definition of the Bochner integral for simple functions),

$$\int_E \boldsymbol{g}(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) = \int_E \sum_{k=1}^{K} \boldsymbol{z}^k \mathbb{1}_{\mathcal{G}_k}(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h})$$

$$\equiv \sum_{k=1}^{K} \boldsymbol{z}^k \mathrm{P}(E \bigcap \mathcal{G}_k) = \begin{bmatrix} \sum_{k=1}^{K} z_{0,1}^k \mathrm{P}(E \bigcap \mathcal{G}_k) \\ \vdots \\ \sum_{k=1}^{K} z_{0,N}^k \mathrm{P}(E \bigcap \mathcal{G}_k) \\ \sum_{k=1}^{K} z_{1,1}^k \mathrm{P}(E \bigcap \mathcal{G}_k) \\ \vdots \end{bmatrix} = \begin{bmatrix} \int_E g_{0,1}(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \\ \vdots \\ \int_E g_{0,N}(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \\ \int_E g_{1,1}(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \\ \vdots \end{bmatrix}.$$

In other words, the desired property that we would like to show for the Bochner integral holds for all simple functions. Using this basic fact, again by definition [29, Definition II.2.1], it holds that, for each $E \in \mathscr{B}(\mathcal{H})$,

$$\int_E \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) = \lim_{m \to \infty} \int_E \boldsymbol{g}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) = \lim_{m \to \infty} \begin{bmatrix} \int_E g_{0,1}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \\ \int_E g_{0,2}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \\ \vdots \\ \int_E g_{0,N}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \\ \int_E g_{1,1}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \\ \vdots \end{bmatrix},$$

where the limit is with respect to the natural norm of $\mathbb{X}$. Therefore, for every $(n, i) \in \mathbb{N} \times \mathbb{N}_N^+$, we have that

$$\begin{aligned} 0 &= \lim_{m \to \infty} \left\| \int_E \boldsymbol{g}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) - \int_E \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \right\|_{\mathbb{X}} \\ &\geq \lim_{m \to \infty} \left| \int_E g_{n,i}^m(\boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) - \left[\int_E \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \right]_{n,i} \right| \\ &= \left| \int_E \lambda_n(\boldsymbol{h}) f_i(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) - \left[\int_E \boldsymbol{\Lambda}_{\mathbb{B}}(\boldsymbol{h}) \otimes \boldsymbol{f}(\boldsymbol{p}(\boldsymbol{h}), \boldsymbol{h}) \mathrm{dP}(\boldsymbol{h}) \right]_{n,i} \right| \geq 0. \end{aligned}$$

Enough said. ∎

Appendix B Detail: Limit Exchange Argument in Section 5.2

For brevity, let

$$\begin{aligned} Z_\alpha(i) &\triangleq f_i(\boldsymbol{p}(\boldsymbol{H}), \boldsymbol{H}), \quad i \in \mathbb{N}_N^+, \\ Z'_\alpha(i) &\triangleq f_i(\boldsymbol{p}'(\boldsymbol{H}), \boldsymbol{H}), \quad i \in \mathbb{N}_N^+ \quad \text{and} \\ Z_\alpha(n, i) &\triangleq f_i(\boldsymbol{p}_\alpha^n(\boldsymbol{H}), \boldsymbol{H}), \quad (n, i) \in \mathbb{N} \times \mathbb{N}_N^+. \end{aligned}$$

We have shown that

$$\begin{gathered} \forall \varepsilon > 0,\ \exists N(\varepsilon),\ \text{such that } \forall n > N(\varepsilon) \text{ and } \forall (m, i) \in \mathbb{N} \times \mathbb{N}_N^+, \\ |\mathbb{E}\{\lambda_m Z_\alpha(n, i)\} - \alpha \mathbb{E}\{\lambda_m Z_\alpha(i)\} - (1 - \alpha) \mathbb{E}\{\lambda_m Z'_\alpha(i)\}| \leq \varepsilon. \end{gathered}$$

In other words, convergence is uniform in m, which also implies that convergence is *uniform over all possible subfamilies* of elements in the countable base $\mathbb{B}$. Equivalently, we can write the statement

$$\begin{gathered} \forall \varepsilon > 0,\ \exists N(\varepsilon),\ \text{such that } \forall n > N(\varepsilon), \\ \sup_i \left[\sup_{\mathbb{F} \subseteq \mathbb{B}} \sup_{\lambda \in \mathbb{F}} \right] |\mathbb{E}\{\lambda Z_\alpha(n, i)\} - \alpha \mathbb{E}\{\lambda Z_\alpha(i)\} - (1 - \alpha) \mathbb{E}\{\lambda Z'_\alpha(i)\}| \leq \varepsilon, \end{gathered}$$

where it holds that

$$\sup_{\mathbb{F}\subseteq\mathbb{B}}\sup_{\lambda\in\mathbb{F}}\bullet=\sup_{\lambda\in\mathbb{B}}\bullet\,.$$

For every choice of $\zeta\in\mathbb{A}_\gamma^i$, $i\in\mathbb{N}_N^+$, as we discuss in Section 5.2, there is a subsequence $\{\lambda_m\}_{m\in\mathcal{K}'_\zeta}$, $\mathcal{K}'_\zeta\subseteq\mathbb{N}$ (corresponding to a subfamily of elements in $\mathbb{B}$) such that

$$\lambda_m\underset{\mathcal{K}'_\zeta\ni m\to\infty}{\longrightarrow}\zeta,\quad \text{P-a.e.}$$

Then, by dominated convergence –note that $Z_\alpha(n,i)$, $Z_\alpha(i)$ and $Z'_\alpha(i)$ are all integrable and all λ_m's are essentially bounded by γ–, we have that

$$\begin{aligned}\mathbb{E}\{\lambda_m Z_\alpha(n,i)\}&\underset{\mathcal{K}'_\zeta\ni m\to\infty}{\longrightarrow}\mathbb{E}\{\zeta Z_\alpha(n,i)\},\\ \mathbb{E}\{\lambda_m Z_\alpha(i)\}&\underset{\mathcal{K}'_\zeta\ni m\to\infty}{\longrightarrow}\mathbb{E}\{\zeta Z_\alpha(i)\}\quad\text{and}\\ \mathbb{E}\{\lambda_m Z'_\alpha(i)\}&\underset{\mathcal{K}'_\zeta\ni m\to\infty}{\longrightarrow}\mathbb{E}\{\zeta Z'_\alpha(i)\}.\end{aligned}$$

In other words, for every $\eta>0$, there is a *common index* $M(n,i,\eta,\zeta)\in\mathcal{K}'_\zeta$, such that, for every $m>M(n,i,\eta,\zeta)$ and in $\mathcal{K}'_\zeta$,

$$\begin{aligned}|\mathbb{E}\{\lambda_m Z_\alpha(n,i)\}-\mathbb{E}\{\zeta Z_\alpha(n,i)\}|&\le\eta,\\ |\mathbb{E}\{\lambda_m Z_\alpha(i)\}-\mathbb{E}\{\zeta Z_\alpha(i)\}|&\le\eta\quad\text{and}\\ |\mathbb{E}\{\lambda_m Z'_\alpha(i)\}-\mathbb{E}\{\zeta Z'_\alpha(i)\}|&\le\eta.\end{aligned}$$

Therefore,

$$\forall\varepsilon>0,\ \exists N(\varepsilon),\ \text{such that }\forall n>N(\varepsilon),\forall i\in\mathbb{N}_N^+,\forall\zeta\in\mathbb{A}_\gamma^i,\forall\eta>0,\text{ and }\forall m>M(n,i,\eta,\zeta)\text{ and in }\mathcal{K}'_\zeta,$$
$$|\mathbb{E}\{\lambda_m Z_\alpha(n,i)\}-\alpha\mathbb{E}\{\lambda_m Z_\alpha(i)\}-(1-\alpha)\mathbb{E}\{\lambda_m Z'_\alpha(i)\}|\le\varepsilon.$$

But under those circumstances, we have

$$\begin{aligned}\varepsilon&\ge|\mathbb{E}\{\lambda_m Z_\alpha(n,i)\}-\alpha\mathbb{E}\{\lambda_m Z_\alpha(i)\}-(1-\alpha)\mathbb{E}\{\lambda_m Z'_\alpha(i)\}|\\ &=|\mathbb{E}\{\lambda_m Z_\alpha(n,i)\}-\mathbb{E}\{\zeta Z_\alpha(n,i)\}+\mathbb{E}\{\zeta Z_\alpha(n,i)\}\\ &\quad-\alpha\mathbb{E}\{\lambda_m Z_\alpha(i)\}+\alpha\mathbb{E}\{\zeta Z_\alpha(i)\}-\alpha\mathbb{E}\{\zeta Z_\alpha(i)\}\\ &\qquad-(1-\alpha)\mathbb{E}\{\lambda_m Z'_\alpha(i)\}+(1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}-(1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}|\\ &\ge\Big||\mathbb{E}\{\zeta Z_\alpha(n,i)\}-\alpha\mathbb{E}\{\zeta Z_\alpha(i)\}-(1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}|\\ &\quad-|\mathbb{E}\{\lambda_m Z_\alpha(n,i)\}-\mathbb{E}\{\zeta Z_\alpha(n,i)\}\\ &\qquad+\alpha(\mathbb{E}\{\zeta Z_\alpha(i)\}-\mathbb{E}\{\lambda_m Z_\alpha(i)\})+(1-\alpha)(\mathbb{E}\{\zeta Z_\alpha(i)\}-\mathbb{E}\{\lambda_m Z_\alpha(i)\})|\Big|\\ &\ge|\mathbb{E}\{\zeta Z_\alpha(n,i)\}-\alpha\mathbb{E}\{\zeta Z_\alpha(i)\}-(1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}|\\ &\quad-|\mathbb{E}\{\lambda_m Z_\alpha(n,i)\}-\mathbb{E}\{\zeta Z_\alpha(n,i)\}\\ &\qquad+\alpha(\mathbb{E}\{\zeta Z_\alpha(i)\}-\mathbb{E}\{\lambda_m Z_\alpha(i)\})+(1-\alpha)(\mathbb{E}\{\zeta Z_\alpha(i)\}-\mathbb{E}\{\lambda_m Z_\alpha(i)\})|,\end{aligned}$$

which implies that

$$|\mathbb{E}\{\zeta Z_\alpha(n,i)\}-\alpha\mathbb{E}\{\zeta Z_\alpha(i)\}-(1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}|\le\varepsilon+\eta+\alpha\eta+(1-\alpha)\eta=\varepsilon+2\eta,$$

and therefore, we have shown that

$$\forall\varepsilon>0,\ \exists N(\varepsilon),\ \text{such that }\forall n>N(\varepsilon),\forall i\in\mathbb{N}_N^+,\forall\zeta\in\mathbb{A}_\gamma^i\text{ and }\forall\eta>0,$$

$$|\mathbb{E}\{\zeta Z_\alpha(n,i)\} - \alpha\mathbb{E}\{\zeta Z_\alpha(i)\} - (1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}| \le \varepsilon + 2\eta.$$

In other words, we can write

$$\begin{gathered}\forall \varepsilon > 0,\ \exists N(\varepsilon),\ \text{such that } \forall n > N(\varepsilon) \text{ and } \forall \eta > 0,\\ \sup_{i\in\mathbb{N}_N^+}\sup_{\zeta\in\mathbb{A}_\gamma^i} |\mathbb{E}\{\zeta Z_\alpha(n,i)\} - \alpha\mathbb{E}\{\zeta Z_\alpha(i)\} - (1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}| \le \varepsilon + 2\eta.\end{gathered}$$

Since η is arbitrary, it is actually true that

$$\begin{gathered}\forall \varepsilon > 0,\ \exists N(\varepsilon),\ \text{such that } \forall n > N(\varepsilon),\\ \sup_{i\in\mathbb{N}_N^+}\sup_{\zeta\in\mathbb{A}_\gamma^i} |\mathbb{E}\{\zeta Z_\alpha(n,i)\} - \alpha\mathbb{E}\{\zeta Z_\alpha(i)\} - (1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}| \le \varepsilon,\end{gathered}$$

and because

$$\begin{aligned}&\sup_{\zeta\in\mathbb{A}_\gamma^i} |\mathbb{E}\{\zeta Z_\alpha(n,i)\} - [\alpha\mathbb{E}\{\zeta Z_\alpha(i)\} + (1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}]| \ge\\ &\qquad\left|\inf_{\zeta\in\mathbb{A}_\gamma^i}\mathbb{E}\{\zeta Z_\alpha(n,i)\} - \inf_{\zeta\in\mathbb{A}_\gamma^i}\alpha\mathbb{E}\{\zeta Z_\alpha(i)\} + (1-\alpha)\mathbb{E}\{\zeta Z'_\alpha(i)\}\right|\end{aligned}$$

we end up with the statement

$$\begin{gathered}\forall \varepsilon > 0,\ \exists N(\varepsilon),\ \text{such that } \forall n > N(\varepsilon),\\ \sup_{i\in\mathbb{N}_N^+}\left|\inf_{\zeta\in\mathbb{A}_\gamma^i}\mathbb{E}\{\zeta Z_\alpha(n,i)\} - \inf_{\zeta\in\mathbb{A}_\gamma^i}\mathbb{E}\{\zeta[\alpha Z_\alpha(i) + (1-\alpha)Z'_\alpha(i)\}\right| \le \varepsilon.\end{gathered}$$

Enough said. ■

Appendix C Strong Duality via Convexity of cl($\mathcal{C}$): Alternative Proof

It is possible to establish strong duality of problem (RCP) by exploiting a characterization of strong duality in general infinite dimensional cone-constrained nonconvex programming by Flores-Bazán and Mastroeni [30, Theorem 3.2]. While the overall argument is indeed elegant, we believe that it is not as transparent and significantly less elementary in comparison with our discussion in Section 5.3. Nevertheless, we find the development valuable.

In what follows, $\mathbb{S}$ is a Hausdorff topological vector space, $\mathbb{Y}$ is a real Hausdorff locally convex topological vector space with topological dual $\mathbb{Y}^*$ and bilinear pairing $\langle\cdot,\cdot\rangle_{\mathbb{Y}^{(*)}}$, $\mathbb{P}\subseteq\mathbb{Y}$ is a nonempty closed convex cone with possibly empty (topological) interior, and $\mathbb{C}$ is a nonempty subset of $\mathbb{S}$. Under this setting and given operators $F:\mathbb{C}\to\mathbb{R}$ and $\boldsymbol{G}:\mathbb{C}\to\mathbb{Y}$, let us consider the infinite-dimensional cone-constrained program

$$\begin{aligned}-\infty < \mathsf{F}^* \triangleq\ &\underset{\boldsymbol{\tau}}{\text{minimize}}\ F(\boldsymbol{\tau})\\ &\text{subject to}\ \boldsymbol{G}(\boldsymbol{\tau}) \in -\mathbb{P}\,,\\ &\qquad\qquad\ \boldsymbol{\tau}\in\mathbb{C}\end{aligned}\tag{CCP}$$

whose feasible set is hereafter assumed to be nonempty; thus $\mathsf{F}^* \in \mathbb{R}$. In direct analogy to the finite-dimensional setting, the Lagrangian dual associated to (CCP) is

$$\sup_{\boldsymbol{\lambda}\in\mathbb{P}^*}\inf_{\boldsymbol{\tau}\in\mathbb{C}} F(\boldsymbol{\tau}) + \langle\boldsymbol{\lambda},\boldsymbol{G}(\boldsymbol{\tau})\rangle_{\mathbb{Y}^{(*)}},$$

where $\mathbb{P}^*$ is the positive polar cone of $\mathbb{P}$. We also define the *image space set* [28]

$$\mathcal{E}_* \triangleq \begin{bmatrix} F(\mathbb{C}) \\ \boldsymbol{G}(\mathbb{C}) \end{bmatrix} - \begin{bmatrix} \mathsf{F}^* \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbb{R}_+ \\ \mathbb{P} \end{bmatrix},$$

where set addition is in the Minkowski sense. Under this general setting, the following powerful result holds, providing necessary and sufficient conditions for strong duality in infinite dimensions.

Theorem 8 (**[30] Characterization of Strong Duality**). *Strong duality holds for (CCP) if and only if*

$$\mathrm{cl}[\mathrm{cone}(\mathrm{conv}(\mathcal{E}_*))] \bigcap \begin{bmatrix} -\mathbb{R}_{++} \\ \{\mathbf{0}\} \end{bmatrix} = \emptyset.$$

It readily follows that the risk-constrained problem (RCP) is an instance of (CCP) and can be expressed as

$$\begin{aligned} -\mathsf{P}^* = \ & \underset{\boldsymbol{x},\boldsymbol{p}(\cdot)}{\text{minimize}} \ -g^o(\boldsymbol{x}) \\ & \text{subject to} \ \begin{bmatrix} \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) + \boldsymbol{x} \\ -\boldsymbol{g}(\boldsymbol{x}) \end{bmatrix} \in -\mathbb{R}_+^{N+N_{\boldsymbol{g}}}, \\ & (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X} \times \Pi \end{aligned} \tag{RCP}$$

with the identifications $\mathbb{S} = \mathbb{R}^N \times \mathcal{L}_1(\mathrm{P},\mathbf{R})$ –assuming for simplicity that $\Pi \subseteq \mathcal{L}_1(\mathrm{P},\mathbf{R})$–, $\mathbb{Y} = \mathbb{R}^{N+N_{\boldsymbol{g}}}$, $\mathbb{P} = \mathbb{R}_+^{N+N_{\boldsymbol{g}}}$, $\mathbb{C} = \mathcal{X} \times \Pi$, and of course $\mathsf{F}^* = -\mathsf{P}^*$. In this case, the set $\mathcal{E}_*$ takes the particular form

$$\begin{aligned} \mathcal{E}_* &= \left\{ (\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \,\middle|\, \begin{matrix} \delta_o = -g^o(\boldsymbol{x}) - (-\mathsf{P}^*) + z_o \\ \boldsymbol{\delta}_r = \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) + \boldsymbol{x} + \boldsymbol{z}_r, \\ \boldsymbol{\delta}_d = -\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{z}_d \end{matrix} \ \text{for some} \ \begin{matrix} (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X} \times \Pi \\ (z_o, \boldsymbol{z}_r, \boldsymbol{z}_d) \in \mathbb{R}_+^{1+N+N_{\boldsymbol{g}}} \end{matrix} \right\} \\ &= \left\{ (\delta_o, \boldsymbol{\delta}_r, \boldsymbol{\delta}_d) \,\middle|\, \begin{matrix} -g^o(\boldsymbol{x}) + \mathsf{P}^* \leq \delta_o \\ \boldsymbol{\rho}(-\boldsymbol{f}(\boldsymbol{p}(\boldsymbol{H}),\boldsymbol{H})) + \boldsymbol{x} \leq \boldsymbol{\delta}_r, \\ -\boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{\delta}_d \end{matrix} \ \text{for some} \ (\boldsymbol{x},\boldsymbol{p}) \in \mathcal{X} \times \Pi \right\}, \end{aligned}$$

or, equivalently,

$$\mathcal{E}_* = -\left[\mathcal{C} - \begin{bmatrix} \mathsf{P}^* \\ \mathbf{0} \end{bmatrix} \right] \triangleq -\mathcal{C}_*.$$

It follows that $\mathcal{E}_*$ and our utility-constraint set $\mathcal{C}$ are essentially equivalent. From Theorem 8 we can readily see that strong duality of (RCP) will follow if we can show that

$$\mathrm{cl}[\mathrm{cone}(\mathrm{conv}(\mathcal{C}_*))] \bigcap \begin{bmatrix} \mathbb{R}_{++} \\ \{\mathbf{0}\} \end{bmatrix} = \emptyset.$$

Next we show that this is indeed the case as a consequence of the convexity of $\mathrm{cl}(\mathcal{C})$, and provided that Slater's constraint qualification holds. First, note that convexity of $\mathrm{cl}(\mathcal{C})$ is equivalent to convexity of $\mathrm{cl}(\mathcal{C}_*)$ (the latter set is just a translation of the former set), and since

$$\begin{aligned} \mathrm{conv}(\mathcal{C}_*) \subseteq \mathrm{conv}(\mathrm{cl}(\mathcal{C}_*)) = \mathrm{cl}(\mathcal{C}_*) \subseteq \mathrm{cone}(\mathrm{cl}(\mathcal{C}_*)) \\ \implies \mathrm{cone}(\mathrm{conv}(\mathcal{C}_*)) \subseteq \mathrm{cone}(\mathrm{cl}(\mathcal{C}_*)), \end{aligned}$$

it will suffice to show that

$$\mathrm{cl}[\mathrm{cone}(\mathrm{cl}(\mathcal{C}_*))] \bigcap \begin{bmatrix} \mathbb{R}_{++} \\ \{\mathbf{0}\} \end{bmatrix} = \emptyset,$$

because $\mathrm{cl}[\mathrm{cone}(\mathrm{cl}(\mathcal{C}_*))] \supseteq \mathrm{cl}[\mathrm{cone}(\mathrm{conv}(\mathcal{C}_*))]$. In other words, if suffices to show that

$$\forall \varepsilon > 0, \quad (\varepsilon, \mathbf{0}, \mathbf{0}) \notin \mathrm{cl}[\mathrm{cone}(\mathrm{cl}(\mathcal{C}_*))].$$

Under Slater's condition, there exist a number $\eta^S > 0$ and vectors $\boldsymbol{\delta}_r^S > \mathbf{0}$ and $\boldsymbol{\delta}_d^S > \mathbf{0}$ such that

$$(-\eta^S, \boldsymbol{\delta}_r^S, \boldsymbol{\delta}_d^S) \in \mathcal{C}_* \subseteq \text{cl}(\mathcal{C}_*) \subseteq \text{cl}[\text{cone}(\text{cl}(\mathcal{C}_*))].$$

Now, suppose that there is an $\varepsilon > 0$ such that $(\varepsilon, \mathbf{0}, \mathbf{0}) \in \text{cl}[\text{cone}(\text{cl}(\mathcal{C}_*))]$. Because of convexity of the latter set, it holds that, for every $\alpha \in [0, 1]$,

$$\boldsymbol{z}_\alpha = (\alpha\varepsilon - (1-\alpha)\eta^S, (1-\alpha)\boldsymbol{\delta}_r^S, (1-\alpha)\boldsymbol{\delta}_d^S) \in \text{cl}[\text{cone}(\text{cl}(\mathcal{C}_*))],$$

and we can choose α such that

$$1 > \alpha > \frac{\eta^S}{\varepsilon + \eta^S} \iff \alpha\varepsilon - (1-\alpha)\eta^S > 0 \quad \Longrightarrow \quad \boldsymbol{z}_\alpha > \mathbf{0}.$$

Since $\boldsymbol{z}_\alpha$ is in the closure of $\text{cone}(\text{cl}(\mathcal{C}_*)) = \{\boldsymbol{x} | \boldsymbol{x} = \beta\boldsymbol{y}, \boldsymbol{y} \in \text{cl}(\mathcal{C}_*), \beta \geq 0\}$ (note that $\text{cl}(\mathcal{C}_*)$ is convex), there exists a sequence of points $\{\boldsymbol{z}_\alpha^n\}_{n\in\mathbb{N}}$ entirely contained in $\text{cone}(\text{cl}(\mathcal{C}_*))$ which converges to $\boldsymbol{z}_\alpha$; that is, each member of such a sequence must be of the form

$$\boldsymbol{z}_\alpha^n = \beta^n \boldsymbol{y}^n, \quad \text{for some } \boldsymbol{y}^n \in \text{cl}(\mathcal{C}_*) \text{ and } \beta^n \geq 0, \quad n \in \mathbb{N}.$$

But because $\boldsymbol{z}_\alpha$ is strictly positive, such a sequence must also be *eventually* strictly positive, which implies the existence of an index n_o such that

$$\mathbf{0} < \boldsymbol{z}_\alpha^{n_o} \iff \boldsymbol{z}_\alpha^{n_o} = \beta^{n_o}\boldsymbol{y}^{n_o}, \quad \text{for some } \mathbf{0} < \boldsymbol{y}^{n_o} \in \text{cl}(\mathcal{C}_*) \text{ and } \beta^{n_o} > 0.$$

This is impossible, because $\text{cl}(\mathcal{C}_*)$ cannot contain strictly positive points; if this could happen, then there would exist another sequence entirely contained in $\mathcal{C}_*$ and converging to that strictly positive point, implying the existence of at least one strictly positive point in $\mathcal{C}_*$, which is absurd –this is essentially the same limiting argument as that used right above. Consequently, the condition of Theorem 8 is verified, confirming that problem (RCP) exhibits strong duality. ■

As a final remark, the reader might notice the similarity of our arguments above with those in the proof of Lemma 5; of course, this is not at all coincidental.

References

[1] Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming: Modeling and Theory, Second Edition, 2nd edn. Society for Industrial and Applied Mathematics, Philadelphia, PA (2014). https://doi.org/10.1137/1.9781611973433

[2] Ribeiro, A.: Optimal Resource Allocation in Wireless Communication and Networking. EURASIP Journal on Wireless Communications and Networking **2012**(1), 272 (2012) https://doi.org/10.1186/1687-1499-2012-272

[3] Boyd, S.P., Vandenberghe, L.: Convex Optimization, p. 716. Cambridge University Press, Cambridge, UK (2004)

[4] Ruszczyński, A.P.: Nonlinear Optimization, p. 448. Princeton University Press, Princeton, NJ, USA (2006). https://doi.org/10.2307/j.ctvcm4hcj

[5] Bertsekas, D.P.: Convex Optimization Theory, 1st edn., p. 246. Athena Scientific, Nashua NH, US (2009). http://www.athenasc.com/convexduality.html

[6] Tassiulas, L., Ephremides, A.: Stability Properties of Constrained Queueing Systems and Scheduling Policies for Maximum Throughput in Multihop Radio Networks. IEEE Transactions on Automatic Control **37**(12), 1936–1948 (1992) https://doi.org/10.1109/9.182479

[7] Neely, M.J., Modiano, E., Rohrs, C.E.: Dynamic Power Allocation and Routing for Time-Varying Wireless Networks. In: IEEE Journal on Selected Areas in Communications, vol. 23, pp. 89–103 (2005). https://doi.org/10.1109/JSAC.2004.837349

[8] Georgiadis, L., Neely, M.J., Tassiulas, L.: Resource Allocation and Cross-Layer Control in Wireless Networks. Foundations and Trends in Networking **1**(1), 1–144 (2006) https://doi.org/10.1561/1300000001

[9] Sidiropoulos, N.D., Davidson, T.N., Zhi-Quan Luo: Transmit Beamforming for Physical-Layer Multicasting. IEEE Transactions on Signal Processing **54**(6), 2239–2251 (2006) https://doi.org/10.1109/TSP.2006.872578

[10] Wei Yu, Lui, R.: Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems. IEEE Transactions on Communications **54**(7), 1310–1322 (2006) https://doi.org/10.1109/TCOMM.2006.877962

[11] Bazerque, J.A., Giannakis, G.B.: Distributed scheduling and resource allocation for cognitive ofdma radios. In: Proceedings of the 2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications, CrownCom, pp. 343–350. IEEE, ??? (2007). https://doi.org/10.1109/CROWNCOM.2007.4549822 . http://ieeexplore.ieee.org/document/4549822/

[12] Luo, Z.Q., Zhang, S.: Dynamic Spectrum Management: Complexity and Duality. IEEE Journal on Selected Topics in Signal Processing **2**(1), 57–73 (2008) https://doi.org/10.1109/JSTSP.2007.914876

[13] Neely, M.J.: Stochastic Network Optimization with Application to Communication and Queueing Systems. Morgan and Claypool Publishers, Switzerland (2010)

[14] Ribeiro, A., Giannakis, G.B.: Separation principles in wireless networking. IEEE Transactions on Information Theory **56**(9), 4488–4505 (2010) https://doi.org/10.1109/TIT.2010.2053897

[15] Ribeiro, A.: Ergodic Stochastic Optimization Algorithms for Wireless Communication and Networking. IEEE Transactions on Signal Processing **58**(12), 6369–6386 (2010) https://doi.org/10.1109/TSP.2010.2057247

[16] Hu, Y., Ribeiro, A.: Optimal Wireless Networks Based on Local Channel State Information. IEEE Transactions on Signal Processing **60**(9), 4913–4929 (2012) https://doi.org/10.1109/TSP.2012.2202110

[17] He, T., Wang, X., Ni, W.: Optimal Chunk-Based Resource Allocation for OFDMA Systems With Multiple BER Requirements. IEEE Transactions on Vehicular Technology **63**(9), 4292–4301 (2014) https://doi.org/10.1109/TVT.2014.2313354

[18] Naderializadeh, N., Avestimehr, A.S.: ITLinQ: A New Approach for Spectrum Sharing in Device-to-Device Communication Systems. IEEE Journal on Selected Areas in Communications **32**(6), 1139–1151 (2014) https://doi.org/10.1109/JSAC.2014.2328102

[19] Eisen, M., Zhang, C., Chamon, L.F.O., Lee, D.D., Ribeiro, A.: Learning Optimal Resource Allocations in Wireless Systems. IEEE Transactions on Signal Processing **67**(10), 2775–2790 (2019) https://doi.org/10.1109/TSP.2019.2908906

[20] Eisen, M., Gatsis, K., Pappas, G.J., Ribeiro, A.: Learning in Wireless Control Systems Over Nonstationary Channels. IEEE Transactions on Signal Processing **67**(5), 1123–1137 (2019) https://doi.org/10.1109/TSP.2018.2890056

[21] Kalogerias, D.S., Eisen, M., Pappas, G.J., Ribeiro, A.: Model-Free Learning of Optimal Ergodic

Policies in Wireless Systems. IEEE Transactions on Signal Processing **68**, 6272–6286 (2020) https://doi.org/10.1109/TSP.2020.3030073

[22] Chamon, L.F.O., Eldar, Y.C., Ribeiro, A.: Functional Nonlinear Sparse Models. IEEE Transactions on Signal Processing **68**, 2449–2463 (2020) https://doi.org/10.1109/TSP.2020.2982834 arXiv:1811.00577

[23] Chamon, L.F.O., Paternain, S., Calvo-Fullana, M., Ribeiro, A.: Constrained Learning with Non-Convex Losses. IEEE Transactions on Information Theory (2022) https://doi.org/10.1109/TIT.2022.3187948 9813433

[24] Balseiro, S.R., Besbes, O., Pizarro, D.: Survey of Dynamic Resource-Constrained Reward Collection Problems: Unified Model and Analysis. Operations Research (2023) https://doi.org/10.1287/opre.2023.2441 . Accessed 2023-07-20

[25] Gatsis, K., Pajic, M., Ribeiro, A., Pappas, G.J.: Opportunistic Control Over Shared Wireless Channels. IEEE Transactions on Automatic Control **60**(12), 3140–3155 (2015) https://doi.org/10.1109/TAC.2015.2416922

[26] Gatsis, K., Ribeiro, A., Pappas, G.J.: Random Access Design for Wireless Control Systems. Automatica **91**, 1–9 (2018) https://doi.org/10.1016/J.AUTOMATICA.2018.01.021

[27] Yu, W., Lui, R.: Dual Methods for Nonconvex Spectrum Optimization of Multicarrier Systems. IEEE Transactions on Communications **54**(7), 1310–1322 (2006) https://doi.org/10.1109/TCOMM.2006.877962

[28] Giannessi, F.: Constrained Optimization and Image Space Analysis. Springer, New York, NY (2005). https://doi.org/10.1007/0-387-28020-0

[29] Diestel, J., Uhl, J.R.: Vector Measures, Math. surv edn. American Mathematical Society, Providence (1977)

[30] Flores-Bazán, F., Mastroeni, G.: Strong Duality in Cone Constrained Nonconvex Optimization. SIAM Journal on Optimization **23**(1), 153–169 (2013) https://doi.org/10.1137/120861400 . Accessed 2022-09-09

[31] Luenberger, D.G.: Optimization by Vector Space Methods. Wiley, New York, NY (1968)

[32] Yaylali, G., Kalogerias, D.S.: Robust and Reliable Stochastic Resource Allocation via Tail Waterfilling. Technical report (May 2023). https://doi.org/10.48550/arXiv.2305.01053 . arXiv:2305.01053. http://arxiv.org/abs/2305.01053 Accessed 2023-07-03

[33] Rockafellar, R.T., Uryasev, S.: Optimization of Conditional Value-at-Risk. Journal of Risk **2**, 21–41 (2000) https://doi.org/10.2307/1165345 arXiv:arXiv:1011.1669v3

[34] Ogryczak, W., Ruszczyński, A.: From Stochastic Dominance to Mean-Risk Models: Semideviations as Risk Measures. European Journal of Operational Research **116**(1), 33–50 (1999) https://doi.org/10.1016/S0377-2217(98)00167-2

[35] Ogryczak, W., Ruszczyński, A.: Dual Stochastic Dominance and Related Mean-Risk Models. SIAM Journal on Optimization **13**(1), 60–78 (2002) https://doi.org/10.1137/S1052623400375075

[36] Balbás, A., Garrido, J., Mayoral, S.: Properties of Distortion Risk Measures. Methodology and Computing in Applied Probability **11**(3), 385–399 (2009) https://doi.org/10.1007/s11009-008-9089-z . Accessed 2023-07-23

[37] Uhl, J.J.: The Range of a Vector-Valued Measure. Proceedings of the American Mathematical Society **23**(1), 158 (1969) https://doi.org/10.2307/2037509

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

[39] Krokhmal, P., Zabarankin, M., Uryasev, S.: Modeling and optimization of risk. Surveys in Operations Research and Management Science **16**(2), 49–66 (2011) https://doi.org/10.1016/j.sorms.2010.08.001 . Accessed 2023-08-22

[40] Knowles, G.: Lyapunov vector measures. SIAM Journal on Control **13**(2), 294–303 (1975) https://doi.org/10.1137/0313017

[41] Kadets, V.M., Schechtman, G.: Lyapunov's theorem for ℓ_p-valued measures. Algebra i Analiz **4**(5), 148–154 (1992). Translation in St. Petersburg Math. J. 4(5), 961–966 (1993)

[42] Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis vol. 317, p. 734. Springer, Heidelberg, DE (2004). https://doi.org/10.1007/978-3-642-02431-3

[43] Aliprantis, C.D., Border, K.: Infinite Dimensional Analysis: A Hitchhiker's Guide. Springer, Heidelberg, DE (2006). https://doi.org/10.1007/3-540-29587-9

[44] Blackwell, D.: The Range of Certain Vector Integrals. Proceedings of the American Mathematical Society **2**(3), 390 (1951) https://doi.org/10.2307/2031763

[45] Strömberg, T.: The operation of infimal convolution. Dissertationes Mathematicae **352** (1996)

[46] Attouch, H., Buttazzo, G., Michaille, G.: Variational Analysis in Sobolev and BV Spaces. Society for Industrial and Applied Mathematics, Philadelphia, PA (2014)

[47] Attouch, H., Brezis, H.: Duality for the sum of convex functions in general Banach spaces. In: Barroso, J.A. (ed.) Aspects of Mathematics and Its Applications. North-Holland Mathematical Library, vol. 34, pp. 125–133. Elsevier, North Holland (1986). https://doi.org/10.1016/S0924-6509(09)70252-1

[48] Mosco, U.: On the continuity of the Young-Fenchel transform. Journal of Mathematical Analysis and Applications **35**(3), 518–535 (1971) https://doi.org/10.1016/0022-247X(71)90200-9

[49] Rockafellar, R.T.: Extension of Fenchel's duality theorem for convex functions. Duke Mathematical Journal **33**(1), 81–89 (1966)

[50] Kouri, D.P., Surowiec, T.M.: Epi-regularization of risk measures. Mathematics of Operations Research **45**(2), 774–795 (2019)

[51] Beer, G.: On mosco convergence of convex sets. Bulletin of the Australian Mathematical Society **38**(2), 239–253 (1988) https://doi.org/10.1017/S0004972700027519

[52] Kalogerias, D.S., Eisen, M., Pappas, G.J., Ribeiro, A.: A zeroth-order learning algorithm for ergodic optimization of wireless systems with no models and no gradients. In: ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2020-May, pp. 5045–5049. Institute of Electrical and Electronics Engineers Inc., Barcelona, ES (2020). https://doi.org/10.1109/ICASSP40776.2020.9053942

[53] Kalogerias, D.S., Powell, W.B.: Recursive Optimization of Convex Risk Measures: Mean-Semideviation Models. arXiv preprint, arXiv:1804.00636 (2018) arXiv:1804.00636