When Wasserstein DRO Reduces Exactly: Complete Characterizations of Projection Equivalence and Regularization

Chenling Huang[1], Jonathan Yu-Meng Li[2] and Tiantian Mao[1]

[1] *Department of Statistics and Finance, School of Management*

University of Science and Technology of China, Hefei, Anhui, China.

[2] *Telfer School of Management, University of Ottawa*

Ottawa, Ontario K1N 6N5, Canada

Emails: hcl0210@mail.ustc.edu.cn; jonathan.li@telfer.uottawa.ca; tmao@ustc.edu.cn

February 6, 2026

Abstract

Wasserstein distributionally robust optimization (DRO), a leading paradigm in data-driven decision-making, requires evaluating worst-case risk over a high-dimensional Wasserstein ball. We study when this worst-case evaluation admits an exact reduction to a one-dimensional formulation, in the sense that it can be carried out over a one-dimensional Wasserstein ball centered at the projected reference distribution. We refer to this property as projection equivalence. We investigate projection equivalence across several classes of risk functionals. Starting from general law-invariant risk functionals and progressing through monotone risk functionals, coherent risk measures, and further specialized subclasses, we provide a complete characterization by giving necessary and sufficient conditions on the loss function under which projection equivalence holds. Beyond simplifying worst-case risk evaluation, our characterization also identifies when the worst-case problem admits an exact regularization reformulation, substantially extending previously known results. Applications to distributionally robust chance-constrained programs and classification problems are presented.

1 Introduction

Wasserstein distributionally robust optimization (DRO) has emerged as a dominant paradigm for optimization under uncertainty, with growing prominence across operations research, statistics, finance, and machine learning. Its strength lies in safeguarding decisions against distributional ambiguity while delivering strong out-of-sample guarantees. In its most general form, Wasserstein DRO can be written as

$$\min_{f\in\mathcal{F}} \sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \rho^F\big(f(\boldsymbol{\xi})\big),$$

where $\mathcal{F}$ denotes the admissible decision class, f represents a decision-dependent loss function, ρ is a risk functional, and $\mathbb{B}_p(F_0, \varepsilon)$ is the p-Wasserstein ball of radius $\varepsilon > 0$ centered at a nominal distribution F_0. Through the choice of ρ, this formulation accommodates a wide array of performance criteria: expectation in classical Wasserstein DRO, risk measures in finance, statistical functionals in inference, and loss- or utility-based objectives in machine learning.

The primary challenge in solving Wasserstein DRO is the inner maximization, which we refer to as the worst-case risk problem:

$$\sup_{F \in \mathbb{B}_p(F_0, \varepsilon)} \rho^F\big(f(\boldsymbol{\xi})\big). \tag{1}$$

When the random vector $\boldsymbol{\xi}$ is high-dimensional, evaluating (1) is often the main bottleneck, and overcoming this difficulty is important for both the theoretical analysis and the practical use of Wasserstein DRO. In some cases, the high-dimensional worst-case problem (1) admits an *exact* reduction to a one-dimensional formulation over a univariate p-Wasserstein ball. We refer to this property as *projection equivalence*:

$$\sup_{F \in \mathbb{B}_p(F_0, \varepsilon)} \rho^F(f(\boldsymbol{\xi})) = \sup_{G \in \mathcal{C}_p(f|F_0, \varepsilon)} \rho^G(X), \tag{2}$$

where $\mathcal{C}_p(f|F_0, \varepsilon)$ denotes the one-dimensional p-Wasserstein ball centered at G_0, the distribution of $f(\boldsymbol{\zeta})$ for $\boldsymbol{\zeta} \sim F_0$. Thus, the high-dimensional worst-case risk evaluation reduces exactly to a one-dimensional counterpart, yielding substantial benefits for both computation and analysis, and in many cases leading to closed-form or efficiently computable solutions.

Projection equivalence (2) has thus far been observed only in limited settings, most notably when the loss function f is linear (or affine). In this case, Mao et al. (2022), Wu et al. (2022) and Aolaritei et al. (2023) obtain projection equivalence by establishing a *set-level equivalence*: the projection of the high-dimensional Wasserstein ball under a linear map coincides exactly with the one-dimensional Wasserstein ball $\mathcal{C}_p(f|F_0, \varepsilon)$. In more general settings, only a set-inclusion relationship can be established: the projection of the high-dimensional Wasserstein ball through f is contained in $\mathcal{C}_p(f|F_0, \varepsilon)$ (see, e.g., Santambrogio (2015)). This in turn implies that the one-dimensional worst-case problem, i.e., the right-hand side of (2), provides at best an upper bound on the full-dimensional worst-case problem, i.e., the left-hand side of (2).

Set-level equivalence is stronger than projection equivalence and may therefore be more restrictive than necessary. For reducing the worst-case risk problem, what matters is equality of the worst-case values, not equality of the ambiguity sets. The linear case is thus only a narrow special instance, leaving open the fundamental question of when exact reduction is possible for broader classes of loss functions and risk functionals, even without set-level equivalence.

To the best of our knowledge, no prior work has provided a complete characterization of when projection equivalence holds beyond the linear setting. More generally, whether such an equivalence holds depends on the class of risk functionals used to evaluate risk. In this paper, we close this gap by developing a hierarchy of results: starting from the most general law-invariant risk functionals and then specializing to monotone functionals, coherent risk measures, and further subclasses, we derive necessary and sufficient conditions on the loss functions f under which projection equivalence (2) holds, thus providing a complete characterization. On the one hand, our results reveal that solving the high-dimensional worst-case risk problem via its one-dimensional counterpart is possible for classes of loss functions that extend far beyond the linear case. On the other hand, and perhaps even more theoretically intriguing, our results also constitute impossibility results: such a reduction is provably not possible for any loss function outside the identified classes. This establishes a sharp boundary for projection equivalence in Wasserstein DRO, delineating precisely when exact reduction is feasible and when it is not. As an application, we show how our reduction results enable an exact reformulation of Wasserstein chance-constrained programs, extending in a nontrivial way the previous results of Xie (2021) and Chen et al. (2024) from the type-1 Wasserstein setting to general type-p.

As another key benefit of reduction, our results show that projection equivalence enables the identification of broader conditions under which Wasserstein DRO problems admit an exact reformulation as regularized optimization problems (Pflug et al. (2012); Blanchet et al. (2019); Shafieezadeh-Abadeh et al. (2019); Gao et al. (2024); Wu et al. (2022)). Such reformulations are of great interest in both optimization and machine learning, as they reveal when Wasserstein DRO can be interpreted and solved through regularization schemes commonly applied in practice. Previously, exact regularization reformulations were known only in restricted cases: in particular, when the risk functional is the expectation (Shafieezadeh-Abadeh et al. (2019); Gao et al. (2024)), or more generally, for other risk functionals but limited to linear loss functions (Wu et al. (2022)). Our results extend these findings substantially by pinpointing precisely when such reformulations exist across broader classes of risk functionals and loss functions.

We further extend the reduction result (2) to the classification setting. We show that exact reduction remains possible, though for a more restricted class of loss functions than in our baseline setting; nonetheless, it still goes well beyond the linear classifiers studied in existing work (e.g., Kuhn et al. (2019); Ho-Nguyen and Wright (2023)).

To provide a high-level view of our main results, including set-level equivalence, projection equivalence, regularization, and the exact characterization of loss functions, we summarize their relationships in Figure 1 for general $p \geqslant 1$ and Figure 2 for $p = 1$.

Our contributions. This paper makes the following advances:

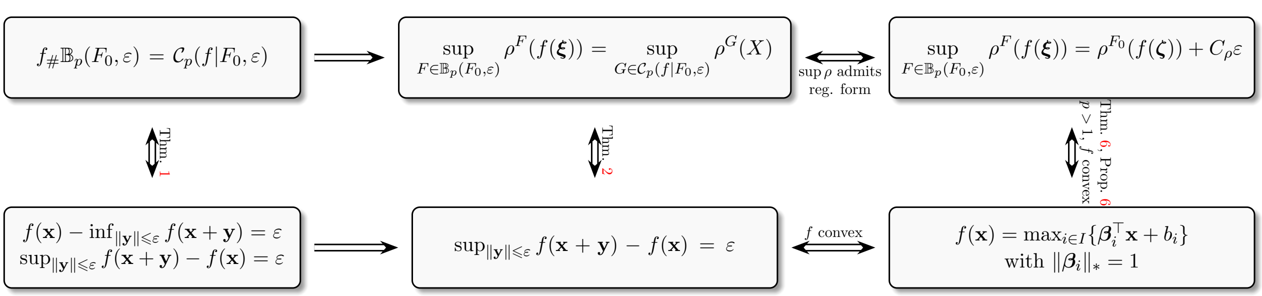
1. **Complete characterization.** We provide necessary and sufficient conditions under which the high-dimensional worst-case risk problem reduces exactly to its one-dimensional counterpart, offering the first full characterization of projection equivalence in Wasserstein DRO.

2. **Beyond set preservation.** We show that projection equivalence can hold even without set-level equivalence, revealing a broader class of (f, ρ) pairs than previously recognized and establishing sharp impossibility boundaries beyond them.

3. **Functional characterization.** For convex losses, we identify the exact family of functions in closed-form admitting projection equivalence, subsuming affine and piecewise-linear forms as special cases.

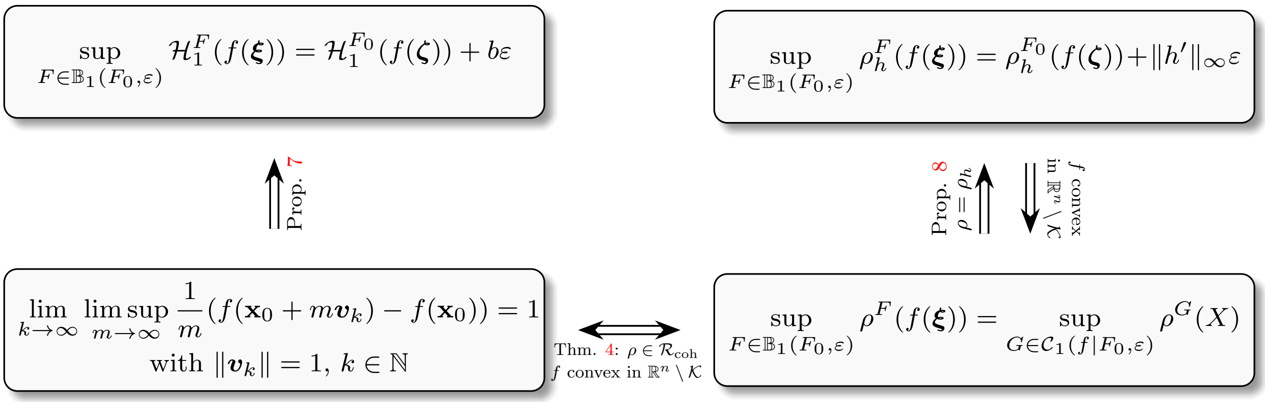
4. **Regularization reformulations.** Leveraging projection equivalence, we derive precise conditions under which Wasserstein DRO admits exact regularization representations, substantially extending prior results beyond expectation and linear losses.

Figure 1: Illustration of the relationships in Theorems 1, 2, 6 and Proposition 6



Notes. The figure is presented for the special case $\mathrm{Lip}(f) = c_f = 1$, for simiplicity. The three conditions in the first layer are assumed to hold for all $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and all $\varepsilon > 0$, whereas the first two conditions in the second layer are assumed to hold for all $\mathbf{x} \in \mathbb{R}^n$ and all $\varepsilon > 0$. The equivalence stated in Theorem 2 holds under the assumption that $\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \rho^F(f(\boldsymbol{\xi})) = \sup_{G\in\mathcal{C}_p(f|F_0,\varepsilon)} \rho^G(X)$ for any monotone risk measure ρ. We say that "$\sup \rho$ admits a regularized form" if there exists a constant $C_\rho \in \mathbb{R}$ such that $\sup_{G\in\mathcal{C}_p(f|F_0,\varepsilon)} \rho^G(X) = \rho^{F_0}(f(\boldsymbol{\zeta})) + C_\rho\varepsilon$. Theorem 6 shows that, when $p > 1$, $\rho = \rho_h$ is a convex distortion risk measure and f is convex, the two rightmost conclusions are equivalent, with $C_\rho = \|h'\|_q$, where q is the Hölder conjugate of p. Proposition 6 shows that, when $p > 1$, $\rho = \mathcal{H}_p$ is defined by (29) with loss $\ell(z,t) = c(z-t)_+$ for $c > 1$ and f is convex, the same equivalence holds with $C_\rho = c$.

Figure 2: Relationships among Theorem 4 and Propositions 7, 8 when $p = 1$



Notes. The figure is presented for the special case $\mathrm{Lip}(f) = c_f = 1$, for simiplicity. The first condition in the second layer (left box) is assumed to hold at some $\mathbf{x}_0 \in \mathbb{R}^n$, while the remaining conditions in all other boxes are assumed to hold for all $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and all $\varepsilon > 0$. Here, $\mathcal{K}$ denotes a bounded set. The functional $\mathcal{H}_1^F$ is defined as $\mathcal{H}_1^F(X) = \inf_{t\in\mathbb{R}} \mathbb{E}^F[t + \ell(X,t)]$, where ℓ satisfies the conditions of Proposition 7, and $\mathrm{Lip}(\ell(\cdot,t)) = b$ for all $t \in \mathbb{R}$. In Proposition 8, ρ_h denotes a convex distortion risk measure.

2 Preliminaries

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be an atomless probability space. A random vector $\boldsymbol{\xi}$ is a measurable mapping from Ω to $\mathbb{R}^n$, $n \in \mathbb{N}$. Denote by $F_{\boldsymbol{\xi}}$ the distribution of $\boldsymbol{\xi}$ under $\mathbb{P}$. Denote by $\mathcal{M}(\mathbb{R}^n)$ the set of all distributions on $\mathbb{R}^n$. For $p \geqslant 1$, let $L^p := L^p(\Omega, \mathcal{A}, \mathbb{P})$ be the set of all random variables with finite pth moment and $\mathcal{M}_p(\mathbb{R}^n)$ be the set of all distributions on $\mathbb{R}^n$ with finite pth moment in each component. For any norm $\|\cdot\|$ on $\mathbb{R}^n$, its dual norm $\|\cdot\|_*$ is defined as $\|\mathbf{y}\|_* = \sup_{\|\mathbf{x}\|\leqslant 1} \mathbf{x}^\top \mathbf{y}$. Let q denote the Hölder conjugate of p, i.e., $1/p + 1/q = 1$. For a real number $x \in \mathbb{R}$, we use $x_+ = \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$; and for $m \in \mathbb{N}$, denote by $[m] = \{1, \ldots, m\}$. Let $\mathbf{e}_i \in \mathbb{R}^n$ be the vector whose ith element is 1 and all other elements are 0 for $i \in [n]$. Let $\mathbf{x}_{-i}$ denote the vector obtained by removing the i-th component from $\mathbf{x} \in \mathbb{R}^n$. Similarly, $\mathbf{x}_{-(i,j)}$ denote the vector obtained by removing the i-th and j-th components. Denote by $\delta_{\mathbf{z}}$ the Dirac distribution at $\mathbf{z} \in \mathbb{R}^n$. We denote by $\mathbf{x} \circ \mathbf{y}$ the Hadamard (element-wise) product of vectors $\mathbf{x}$ and $\mathbf{y}$, i.e., the vector whose i-th component is given by $(\mathbf{x} \circ \mathbf{y})_i = x_i y_i$.

For any two n-dimensional distributions F_1 and F_2 on $\mathcal{M}_p(\mathbb{R}^n)$, the type-$p$ Wasserstein metric is defined as

$$\mathrm{W}_p(F_1, F_2) := \inf_{\pi\in\Pi(F_1,F_2)} \left(\mathbb{E}^\pi[\|\boldsymbol{\xi}_1 - \boldsymbol{\xi}_2\|^p]\right)^{1/p}, \tag{3}$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^n$, and $\Pi(F_1, F_2)$ denotes the set of all distributions on $\mathbb{R}^n \times \mathbb{R}^n$ with marginals F_1 and F_2. We define the ball of distributions $\mathbb{B}_p(F_0, \varepsilon)$ on $\mathbb{R}^n$ as

$$\mathbb{B}_p(F_0, \varepsilon) = \{F \in \mathcal{M}_p(\mathbb{R}^n) : \mathrm{W}_p(F, F_0) \leqslant \varepsilon\}, \tag{4}$$

and refer to it as the type-p Wasserstein ball throughout this paper.

We begin by formalizing the notion of a *projection-induced ambiguity set*, along with the classes of risk functionals under which our main results are developed. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a loss function, $F_0 \in \mathcal{M}_p(\mathbb{R}^n)$ be a nominal distribution, and $\varepsilon > 0$. The *projection-induced p-Wasserstein ball* is defined as

$$f_{\#}\mathbb{B}_p(F_0, \varepsilon) := \left\{ F_{f(\boldsymbol{\xi})} : F_{\boldsymbol{\xi}} \in \mathbb{B}_p(F_0, \varepsilon) \right\},$$

where $F_{f(\boldsymbol{\xi})}$ denotes the distribution of the scalar random variable $f(\boldsymbol{\xi})$ for $\boldsymbol{\xi} \sim F_{\boldsymbol{\xi}}$. We also write

$$\mathcal{C}_p(f \mid F_0, \varepsilon) := \mathbb{B}_p(G_0, \varepsilon) \subseteq \mathcal{M}_p(\mathbb{R})$$

for the *one-dimensional* type-p Wasserstein ball centered at G_0, the distribution of $f(\boldsymbol{\zeta})$ for $\boldsymbol{\zeta} \sim F_0$. If f is Lipschitz continuous, we define its Lipschitz constant with respect to the norm used in the Wasserstein metric as

$$\mathrm{Lip}(f) := \sup_{\mathbf{x} \neq \mathbf{y}} \frac{|f(\mathbf{x}) - f(\mathbf{y})|}{\|\mathbf{x} - \mathbf{y}\|}.$$

Let $\mathcal{X}$ denote a space of real-valued random variables. We refer to any functional $\rho : \mathcal{X} \to \mathbb{R}$ as a *risk functional*, and consider the following nested classes:

(i) Law-invariant finite-valued risk functionals. A risk functional ρ is *law-invariant* if $\rho(X) = \rho(Y)$ for all $X, Y \in \mathcal{X}$ such that $X \overset{d}{=} Y$. We assume that $\rho(X) < \infty$ for all $X \in \mathcal{X}$.

(ii) Law-invariant monotone risk functionals. A risk functional ρ is *monotone* if $\rho(X) \leqslant \rho(Y)$ whenever $X \leqslant Y$ almost surely.

(iii) Law-invariant coherent risk measures. A risk functional ρ is *coherent* if it is monotone and satisfies the following properties:

- **Translation invariance:** $\rho(X + m) = \rho(X) + m$ for all $m \in \mathbb{R}$,
- **Positive homogeneity:** $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda \geqslant 0$,
- **Subadditivity:** $\rho(X + Y) \leqslant \rho(X) + \rho(Y)$.

It is well known that any law-invariant, coherent, and lower semicontinuous risk measure $\rho : L^p \to \mathbb{R}$ admits a Kusuoka representation (Kusuoka (2001); Filipovic and Svindland (2007); Shapiro

(2013)) of the form

$$\rho^F(X) = \sup_{\mu \in \mathscr{M}_\rho} \int_0^1 \mathrm{CVaR}_\alpha^F(X)\, \mathrm{d}\mu(\alpha), \tag{5}$$

where $\mathscr{M}_\rho$ is a set of probability measures on $[0,1]$, and $\mathrm{CVaR}_\alpha^F(X)$ is the Conditional Value-at-Risk (CVaR, also called Expected Shortfall, ES) at level $\alpha \in [0,1]$ defined as

$$\mathrm{CVaR}_\alpha^F(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{VaR}_s^F(X)\mathrm{d}s,\ \ \alpha \in [0,1), \quad \text{and} \quad \mathrm{CVaR}_1^F(X) = \mathrm{VaR}_1^F(X),$$

with VaR_α being the Value-at-Risk (VaR) at level $\alpha \in [0,1]$ defined as

$$\mathrm{VaR}_\alpha^F(X) = \inf\{x : F(x) > \alpha\},\ \ \alpha \in [0,1) \ \ \text{and} \ \ \mathrm{VaR}_1^F(X) = \inf\{x : F(x) \geqslant 1\}.$$

In this paper, we call a coherent risk measure ρ *regular* if it admits a Kusuoka representation of the form (5) with [1]

$$C_\rho := \sup_{\mu \in \mathscr{M}_\rho} \int_0^1 \frac{1}{1-\alpha}\, \mathrm{d}\mu(\alpha) < \infty.$$

We denote by $\mathcal{R}_{\mathrm{law}}$, $\mathcal{R}_{\mathrm{mon}}$, and $\mathcal{R}_{\mathrm{coh}}$ the respective classes of law-invariant finite-valued risk functionals, monotone risk functionals, and regular coherent risk measures, where

$$\mathcal{R}_{\mathrm{coh}} \subseteq \mathcal{R}_{\mathrm{mon}} \subseteq \mathcal{R}_{\mathrm{law}}.$$

3 Projection Equivalence for General Risk Functionals

We now present our main results: a complete characterization of the loss functions f for which projection equivalence (2) holds. We proceed in a nested manner, starting from the largest class of law-invariant risk functionals $\mathcal{R}_{\mathrm{law}}$ and then restricting to the more structured classes of monotone and coherent risk measures, $\mathcal{R}_{\mathrm{mon}}$ and $\mathcal{R}_{\mathrm{coh}}$. This refinement reveals progressively broader families of losses f for which projection equivalence can be guaranteed.

3.1 Law-invariant Risk Functionals $\mathcal{R}_{\mathrm{law}}$

We first show that requiring projection equivalence to hold uniformly over all law-invariant risk functionals is essentially as strong as requiring the associated one-dimensional ambiguity sets to coincide, namely set-level equivalence. Theorem 1 establishes necessary and sufficient conditions under which this equivalence holds and provides a complete characterization of the admissible loss

[1] The condition $C_\rho < \infty$ is a standard assumption ensuring that the risk measure ρ takes finite value on L^1.

functions.

Theorem 1. *For $p \geqslant 1$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Then the following statements are equivalent.*

(i) There exists $c_f \geqslant 0$ such that

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F\left(f(\boldsymbol{\xi})\right) = \sup_{G \in \mathcal{C}_p(f|F_0,c_f\varepsilon)} \rho^G(X) \tag{6}$$

holds for any $\rho \in \mathcal{R}_{\text{law}}$, $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$.

(ii) There exists $c_f \geqslant 0$ such that for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$, it holds that

$$f_{\#}\mathbb{B}_p(F_0, \varepsilon) = \mathcal{C}_p(f|F_0, c_f\varepsilon). \tag{7}$$

(iii) The function f is Lipschitz continuous, and satisfies for any $\mathbf{x} \in \mathbb{R}^n$, $\varepsilon > 0$,

$$f(\mathbf{x}) - \inf_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x}+\mathbf{y}) = \sup_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) = \mathrm{Lip}(f)\varepsilon. \tag{8}$$

Set-level equivalence (7) immediately implies projection equivalence (6) by definition. The converse—namely, $(i) \Rightarrow (ii)$—is considerably more delicate. As we show later, this implication is specific to the law-invariant class and need not persist for more refined families of risk measures, for which projection equivalence may hold even in the absence of set-level coincidence.

Set-level equivalence (7) was previously established only for linear losses f (see, e.g., Wu et al. (2022)). Characterization (iii) shows that linearity is not required and, more importantly, enables us to identify a substantially richer class of loss functions for which projection equivalence holds. To highlight the additional flexibility afforded by (iii), we provide two explicit families of admissible losses under the choice of ℓ_1-norm $\|\cdot\| = \|\cdot\|_1$.

Proposition 1. *Let $\|\cdot\| = \|\cdot\|_1$. The following two families of loss functions satisfy condition (iii) in Theorem 1.*

(a) $f(\mathbf{x}) = c(\boldsymbol{\beta}^\top \mathbf{x} + g(\boldsymbol{\eta} \circ \mathbf{x}))$ where $c > 0$, $\|\boldsymbol{\beta}\|_\infty = 1$, $\beta_i, \eta_i \in \{1, -1, 0\}$, $i \in [n]$, $\boldsymbol{\beta} \circ \boldsymbol{\eta} = \mathbf{0}$, and g is a Lipschitz function with $\mathrm{Lip}(g) \leqslant 1$.

(b) $f(\mathbf{x}) = c(|\boldsymbol{\beta}^\top \mathbf{x}| - |\boldsymbol{\nu}^\top \mathbf{x}| + g(\boldsymbol{\eta} \circ \mathbf{x}))$, where $c > 0$, $\|\boldsymbol{\beta}\|_\infty = \|\boldsymbol{\nu}\|_\infty = 1$, $\beta_i, \nu_i, \eta_i \in \{1, -1, 0\}$, $i \in [n]$, $\boldsymbol{\beta} \circ \boldsymbol{\eta} = \boldsymbol{\nu} \circ \boldsymbol{\eta} = \boldsymbol{\beta} \circ \boldsymbol{\nu} = \mathbf{0}$, and g is a Lipschitz function with $\mathrm{Lip}(g) \leqslant 1$.

Notably, the admissible families identified above are fairly broad in the $\|\cdot\|_1$ setting. This breadth is nontrivial: as we show next, once one moves away from $\|\cdot\|_1$, exact reduction can become dramatically more restrictive, and for the $\|\cdot\|_a$ norm with $a \in (1,\infty)$ it is impossible beyond linear loss functions. We provide a refined characterization of functions with $\mathrm{Lip}(f) > 0$ that satisfy (8) under *strictly convex* norms (Clarkson (1936)), i.e., norms for which $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$ and $\|\mathbf{x}-\mathbf{y}\| \neq 0$ imply $\|\mathbf{x}+\mathbf{y}\| < 2$. It is well known that $\|\cdot\|_a$ is strictly convex for $a \in (1,\infty)$. Perhaps surprisingly, under any such norm, the admissible class collapses: (8) holds only for linear f.

Proposition 2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz function with $\mathrm{Lip}(f) > 0$. If $\|\cdot\|$ is a strictly convex norm, then f satisfies (8) if and only if there exists $\boldsymbol{v}$ with $\|\boldsymbol{v}\| = 1$ such that*

$$f(\mathbf{x} + t\boldsymbol{v}) - f(\mathbf{x}) = \mathrm{Lip}(f)t, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\, t \in \mathbb{R}. \tag{9}$$

In particular, if $\|\cdot\| = \|\cdot\|_a$, $a \in (1,\infty)$, then f satisfies (9) if and only if $f(\mathbf{x}) = \boldsymbol{\beta}^\top\mathbf{x} + b$ for some $\boldsymbol{\beta} \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

3.2 Monotone Risk Functionals $\mathcal{R}_{\mathbf{mon}}$

We highlight in this section that projection equivalence can hold for a substantially richer class of loss functions f than in the previous setting (e.g., those in Proposition 1) once we restrict attention to monotone risk functionals. In particular, for this monotone class, projection equivalence no longer requires set-level equivalence (7). Accordingly, we characterize the admissible loss functions through a strictly weaker condition on f.

Theorem 2. *For $p \geqslant 1$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. The following statements are equivalent.*

(i) There exists $c_f \geqslant 0$ such that

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F\left(f(\boldsymbol{\xi})\right) = \sup_{G \in \mathcal{C}_p(f|F_0,c_f\varepsilon)} \rho^G(X) \tag{10}$$

holds for any $\rho \in \mathcal{R}_{\mathrm{mon}}$, $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$.

(ii) The function f is Lipschitz continuous and satisfies

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) = \mathrm{Lip}(f)\varepsilon, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\, \varepsilon > 0. \tag{11}$$

Comparing the characterization (8) in Theorem 1 with (11) above, we see that (8) applies only to functions that are unbounded both above and below, whereas (11) can also apply to functions that are unbounded only from above. Below we provide several examples that fall outside the scope of Theorem 1 but are covered by Theorem 2. The first three are convex, whereas the fourth is not.

Example 1. (i) The norm function: $f(\mathbf{x}) = \|\mathbf{x}\|$. Note that we can rewrite $f(\mathbf{x}) = \sup_{\|\boldsymbol{\beta}\|_*=1} \boldsymbol{\beta}^\top \mathbf{x}$. We have $c_f = 1$ and by Theorem 2, for any $\rho \in \mathcal{R}_{\mathrm{mon}}$, $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$, it holds that

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F (\|\boldsymbol{\xi}\|) = \sup_{G \in \mathbb{B}_p(G_0,\varepsilon)} \rho^G(X),$$

where $G_0 \in \mathcal{M}(\mathbb{R})$ is the distribution of $\|\boldsymbol{\zeta}\|$ and $\boldsymbol{\zeta} \sim F_0$.

(ii) The absolute value linear function: $f(\mathbf{x}) = |\boldsymbol{\beta}^\top \mathbf{x} + b|$, $\boldsymbol{\beta} \in \mathbb{R}^n$, $b \in \mathbb{R}$. We have $c_f = \|\boldsymbol{\beta}\|_*$ and by Theorem 2, for any $\rho \in \mathcal{R}_{\mathrm{mon}}$, $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$, it holds that

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F \left(|\boldsymbol{\beta}^\top \boldsymbol{\xi} + b|\right) = \sup_{G \in \mathbb{B}_p(G_0,\|\boldsymbol{\beta}\|_*\varepsilon)} \rho^G(X),$$

where $G_0 \in \mathcal{M}(\mathbb{R})$ is the distribution of $|\boldsymbol{\beta}^\top \boldsymbol{\zeta} + b|$ and $\boldsymbol{\zeta} \sim F_0$.

(iii) If $\|\cdot\| = \|\cdot\|_1$ is the ℓ_1-norm, then $f(\mathbf{x}) = \max_{i \in I}(\beta_{i1}x_1 + \cdots + \beta_{in}x_n + b_i)$ with $\max_{j\in[n]} |\beta_{ij}| = c$ for $i \in I$, satisfies (11) with $\mathrm{Lip}(f) = c$.

(iv) Take $\mathbb{R}^2$ and the norm $\|(x_1, x_2)\| = \sqrt{x_1^2 + x_2^2}$, $(x_1, x_2) \in \mathbb{R}^2$. Define

$$f(x_1, x_2) = \begin{cases} \max\left\{\frac{-x_1+2x_2}{\sqrt{5}}, \frac{-x_1-2x_2}{\sqrt{5}}\right\}, & x_1 \geqslant 0, \\ \max\left\{\frac{x_1+2x_2}{\sqrt{5}}, \frac{x_1-2x_2}{\sqrt{5}}\right\}, & x_1 < 0. \end{cases}$$

The function f satisfies (11) with $\mathrm{Lip}(f) = 1$.

The first three examples are standard convex losses. This naturally raises the question of whether other convex loss functions also enjoy projection equivalence. To address this question, we establish a representation theorem that provides a complete closed-form characterization of all convex losses admitting projection equivalence.

Proposition 3. *If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $p \geqslant 1$, then the function f satisfying (11) must admit a representation of the form*

$$f(\mathbf{x}) = \max_{i \in I}\{c_f \boldsymbol{\beta}_i^\top \mathbf{x} + b_i\}, \tag{12}$$

where $c_f \geqslant 0$, $\boldsymbol{\beta}_i \in \mathbb{R}^n$, $i \in I$, with $\|\boldsymbol{\beta}_i\|_* = 1$ and $b_i \in \mathbb{R}$.

Remark 1. It is worth noting that Proposition 3 remains valid without convexity when $n = 1$, whereas convexity becomes necessary when $n \geqslant 2$. In particular, Example 1(iv) provides a counterexample. One can verify that this function f satisfies (11), yet it is not convex along the line $x_2 = 0$ and therefore cannot admit the representation (12).

Thus far, we have identified (11) as a necessary condition for projection equivalence to hold uniformly over monotone risk functionals. We next establish a somewhat unexpected result: (11) remains necessary even if projection equivalence is required to hold only for a single monotone risk measure—namely, Value-at-Risk.

Proposition 4. *For $\alpha \in [0,1)$ and $p \geqslant 1$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. Then there exists $c_f \geqslant 0$ such that*

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \mathrm{VaR}_\alpha^F\left(f(\boldsymbol{\xi})\right) = \sup_{G \in \mathcal{C}_p(f|F_0,c_f\varepsilon)} \mathrm{VaR}_\alpha^G(X) \tag{13}$$

holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$ if and only if f satisfies (11).

The above result highlights that it is difficult to expect projection equivalence beyond the class of loss functions characterized by (11). As we show throughout the remainder of the paper, exceptions do exist (see, e.g., Section 3.3), but (11) appears to be fairly tight in general.

Remark 2. It is worth noting that Proposition 4 does not extend to $\alpha = 1$. Indeed, for any f and any $p \in [1,\infty)$, we always have $\sup_{G \in \mathcal{C}_p(f|F_0,c_f\varepsilon)} \mathrm{VaR}_1^G(X) = \infty$. If f is unbounded from above, then $\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \mathrm{VaR}_1^F\left(f(\boldsymbol{\xi})\right) = \infty$, which implies that (13) always holds for $\alpha = 1$ in this case. Therefore, we have Proposition 4 does not hold for $\alpha = 1$ in general.

Before proceeding further, we highlight how Proposition 4 together with Proposition 3 facilitate the solution of important OR/MS problems. Observe that the function $f(\boldsymbol{\xi}) := \max_{i\in I}\{\boldsymbol{\beta}_i^\top \boldsymbol{\xi}/\|\boldsymbol{\beta}_i\|_*\}$ is a special case of (12). By Proposition 4, this yields

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \mathrm{VaR}_\alpha^F\left(\max_{i\in I} \frac{\boldsymbol{\beta}_i^\top \boldsymbol{\xi}}{\|\boldsymbol{\beta}_i\|_*}\right) = \sup_{G \in \mathbb{B}_p(G_0,\varepsilon)} \mathrm{VaR}_\alpha^G(X), \tag{14}$$

where $G_0 \in \mathcal{M}(\mathbb{R})$ is the distribution of $f(\boldsymbol{\zeta})$ and $\boldsymbol{\zeta} \sim F_0$.

We present two important applications of (14). In each case, the reduction to the one-dimensional worst-case Value-at-Risk problem (the right-hand side of (14)) allows us to invoke the CVaR reformulation in Lemma 1 to solve the original high-dimensional problem (the left-hand side of (14)).

Lemma 1. *For $\alpha \in [0,1]$ and $p \geqslant 1$, the worst-case* VaR (14) *is the unique $x \in \mathbb{R}$ satisfying*

$$\frac{\varepsilon^p}{1-\alpha} + \mathrm{CVaR}_\alpha^{F_0}\left(-\left(x - \max_{i\in I} \frac{\boldsymbol{\beta}_i^\top \boldsymbol{\zeta}}{\|\boldsymbol{\beta}_i\|_*} \right)_+^p \right) = 0. \tag{15}$$

Example 2 (Worst-case risk over multiple portfolios)**.** Let $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m \in \mathbb{R}^n$ denote m portfolio weight vectors such that all of them share the same dual norm, i.e., $\|\boldsymbol{\beta}_j\|_* = \|\boldsymbol{\beta}_1\|_*$ for all $j \in [m]$. The problem of evaluating the worst-case value-at-risk of the poorest-performing portfolio is

$$\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \mathrm{VaR}_\alpha^F \left(\max_{i\in[m]} \boldsymbol{\beta}_i^\top \boldsymbol{\xi} \right).$$

By applying (14) together with Lemma 1, the problem can be reformulated as the following optimization problem:

$$\min\ x \quad \text{s.t.}\ \ \frac{\varepsilon^p}{1-\alpha}\|\boldsymbol{\beta}_1\|_*^p + \mathrm{CVaR}_\alpha^{F_0}\left(-\left[x - \max_{i\in[m]} \boldsymbol{\beta}_i^\top \boldsymbol{\zeta} \right]_+^p \right) \leqslant 0. \tag{16}$$

This formulation generalizes the result of Chen and Xie (2021), which is restricted to the case $p = 1$, to all $p \geqslant 1$.

Example 3 (Distributionally robust chance-constrainted program)**.** One important application of Wasserstein DRO is distributionally robust chance-constrained programs (Wasserstein DRCCPs), which aim to ensure that constraints hold with high probability under distributional uncertainty. A typical Wasserstein DRCCP takes the form

$$\min_{\mathbf{x}\in S} \quad \mathbf{c}^\top \mathbf{x}, \tag{17}$$

$$\text{s.t.} \quad \inf_{F\in\mathbb{B}_p(F_0,\varepsilon)} \mathbb{P}^F \Big(a_i(\mathbf{x})^\top \boldsymbol{\xi} \leqslant b_i(\mathbf{x}),\ \forall\ i \in [m] \Big) \geqslant 1-\eta, \tag{18}$$

where $S \subseteq \mathbb{R}^k$ is a feasible set of the decision vector $\mathbf{x}$, the vector $\mathbf{c} \in \mathbb{R}^k$ denotes the objective function coefficients, $\{\boldsymbol{y} : a_i(\mathbf{x})^\top \boldsymbol{y} \leqslant b_i(\mathbf{x}),\ \forall\ i \in [m]\} \subseteq \mathbb{R}^n$ is referred to as a safety set for each $\mathbf{x} \in S$, and $\eta \in (0,1)$. For $p = 1$, the feasible set has been reformulated in terms of CVaR by Xie (2021) and Chen et al. (2024). Here, we provide a new perspective and extend the result to general

$p \geqslant 1$. First, the constraint (18) is equivalent to [2]

$$\sup_{F\in\mathbb{B}_p(F_0,\delta)} \mathrm{VaR}^{L,F}_{1-\eta}\left(\max_{i\in[m]}\left\{\frac{a_i(\mathbf{x})^\top\boldsymbol{\xi}-b_i(\mathbf{x})}{\|a_i(\mathbf{x})\|_*}\right\}\right)\leqslant 0,$$

where VaR^L denotes the left-continuous VaR defined by $\mathrm{VaR}^L_\alpha(X)=\inf\{x:\mathbb{P}(X\leqslant x)\geqslant\alpha\}$ for $\alpha\in(0,1)$. One can easily verify Lemma 1 still holds true if we replace VaR_α by the left-continuous VaR^L_α. Applying (14) together with Lemma 1, problem (17) can be reformulated as

$$\begin{aligned}\min_{\mathbf{x}\in S}\ & \mathbf{c}^\top\mathbf{x}\\ \text{s.t.}\ & -\mathrm{CVaR}^{F_0}_{1-\eta}\left[-f(\mathbf{x},\boldsymbol{\zeta})^p\right]\geqslant\frac{\varepsilon^p}{\eta},\quad i\in I(\mathbf{x}),\\ & a_i(\mathbf{x})=0,\ b_i(\mathbf{x})\geqslant 0,\quad i\notin I(\mathbf{x}),\end{aligned}$$

where $f(\mathbf{x},\boldsymbol{\zeta})=\min_{i\in I(\mathbf{x})}\frac{\left(b_i(\mathbf{x})-a_i(\mathbf{x})^\top\boldsymbol{\zeta}\right)_+}{\|a_i(\mathbf{x})\|_*}$ and $I(\mathbf{x})=\{i\in[m]:a_i(\mathbf{x})\neq 0\}$.

Moreover, observing that

$$\min_{i\in I(\mathbf{x})}-\frac{1}{\|a_i(\mathbf{x})\|_*^p}\mathrm{CVaR}^{F_0}_{1-\eta}\left[-(b_i(\mathbf{x})-a_i(\mathbf{x})^\top\boldsymbol{\xi})_+^p\right]\geqslant-\mathrm{CVaR}^{F_0}_{1-\eta}\left[-f(\mathbf{x},\boldsymbol{\zeta})^p\right],$$

we obtain the following tractable optimization problem, which provides a conservative approximation of (17):

$$\begin{aligned}\min_{\mathbf{x}\in S}\ & \mathbf{c}^\top\mathbf{x}\\ \text{s.t.}\ & -\frac{1}{\|a_i(\mathbf{x})\|_*^p}\mathrm{CVaR}^{F_0}_{1-\eta}\left[-(b_i(\mathbf{x})-a_i(\mathbf{x})^\top\xi_i)_+^p\right]\geqslant\frac{\varepsilon^p}{\eta},\ i\in I(\mathbf{x}),\\ & a_i(\mathbf{x})=0,\ b_i(\mathbf{x})\geqslant 0,\quad i\notin I(\mathbf{x}).\end{aligned}$$

This upper-bound formulation generalizes the result of Xie (2021), which is restricted to $p=1$, to arbitrary $p\geqslant 1$.

3.3 Coherent Risk Functionals $\mathcal{R}_{\mathrm{coh}}$

Finally, we study projection equivalence for the coherent subclass of monotone, law-invariant risk measures, focusing on those admitting a Kusuoka representation. It turns out that in this

[2]For a random variable X with distribution F and $\alpha\in(0,1)$, it holds that $F(0)\geqslant\alpha$ if and only if $\mathrm{VaR}^L_\alpha(X)\leqslant 0$. To see this, first assume that $F(0)\geqslant\alpha$. By definition of VaR, we have $\mathrm{VaR}^L_\alpha(X)=\inf\{x:F(x)\geqslant\alpha\}\leqslant 0$ as $0\in\{x:F(x)\geqslant\alpha\}$. Next assume that $\mathrm{VaR}^L_\alpha(X)\leqslant 0$, that is, $\inf\{x:F(x)\geqslant\alpha\}\leqslant 0$. There exist $x_n\downarrow 0$ as $n\to\infty$ such that $F(x_n)\geqslant\alpha$. By right-continuity of F, we have $F(0)\geqslant\alpha$.

setting the choice of Wasserstein ball—specifically, the order p—is decisive: the regimes $p > 1$ and $p = 1$ behave fundamentally differently. We first show that when $p > 1$, the loss-function condition (11) identified in the preceding section remains necessary, even after restricting attention to coherent risk measures.

Theorem 3. *For $p > 1$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. There exists $c_f \geqslant 0$ such that*

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F(f(\boldsymbol{\xi})) = \sup_{G \in \mathcal{C}_p(f|F_0,c_f\varepsilon)} \rho^G(X) \tag{19}$$

holds for any $\rho \in \mathcal{R}_{\mathrm{coh}}$, $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$ if and only if f satisfies (11).

Similar to Proposition 4, (11) remains necessary even if projection equivalence is required to hold only for a single coherent risk measure—namely, expectation.

Proposition 5. *For $p > 1$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. There exists $c_f \geqslant 0$ such that*

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \mathbb{E}^F\left[f(\boldsymbol{\xi})\right] = \sup_{G \in \mathcal{C}_p(f|F_0,c_f\varepsilon)} \mathbb{E}^G[X]$$

holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$ if and only if f satisfies (11).

In sharp contrast, the case $p = 1$ is markedly different: under the type-1 Wasserstein ball, projection equivalence can hold for a significantly broader class of loss functions. We provide a complete characterization of this class.

Theorem 4. *Let $\rho \in \mathcal{R}_{\mathrm{coh}}$ and $f : \mathbb{R}^n \to \mathbb{R}$ be a function. The following statements hold.*

(i) If f is Lipschitz continuous and there exist $\mathbf{x}_0 \in \mathbb{R}^n$ and $\boldsymbol{v}_k \in \mathbb{R}^n$ with $\|\boldsymbol{v}_k\| = 1, k \in \mathbb{N}$ such that

$$\lim_{k\to\infty} \limsup_{m\to\infty} \frac{1}{m}(f(\mathbf{x}_0 + m\boldsymbol{v}_k) - f(\mathbf{x}_0)) = \mathrm{Lip}(f), \tag{20}$$

then for any $F_0 \in \mathcal{M}\left(\mathbb{R}^n\right)$ and $\varepsilon > 0$, it holds with $c_f = \mathrm{Lip}(f)$ that

$$\sup_{F \in \mathbb{B}_1(F_0,\varepsilon)} \rho^F\left(f\left(\boldsymbol{\xi}\right)\right) = \sup_{G \in \mathcal{C}_1(f|F_0,\, c_f\varepsilon)} \rho^G(X). \tag{21}$$

(ii) If there exists a bounded set $\mathcal{K} \subseteq \mathbb{R}^n$ such that f coincides with some convex function on $\mathbb{R}^n \backslash \mathcal{K}$, then we have (20) is also necessary for (21) to hold for any $F_0 \in \mathcal{M}\left(\mathbb{R}^n\right)$ and $\varepsilon > 0$.

(iii) If f is convex, then there exists $c_f \geqslant 0$ such that (21) holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$ if and only if f is Lipschitz continuous.

Conditions (ii) and (iii) establish necessity of the characterization (20) for, respectively, potentially nonconvex and convex loss functions. In particular, condition (ii) shows that (20) is necessary for any loss function with a "convex tail". We conclude this section by presenting examples of loss functions that satisfy (20).

Example 4. The following are examples of Lipschitz continuous functions that satisfy (20).

(i) Norm-based check loss: $f(\mathbf{x}) = \big(\tau - \mathbb{1}_{\{\|\mathbf{x}\| \leqslant c\}}\big)(\|\mathbf{x}\| - c)$, where $\tau \geqslant 1/2$ and $c \geqslant 0$. Then $\mathrm{Lip}(f) = \tau$. One can verify that $f(\mathbf{x}) = \tau(\|\mathbf{x}\| - c)$ on $\mathbb{R}^n \setminus \mathcal{K}$ with $\mathcal{K} := \{\mathbf{x} : \|\mathbf{x}\| \leqslant c\}$.

(ii) Huber check loss: $f(\mathbf{x}) = \big(\tau - \mathbb{1}_{\{g_\alpha(\|\mathbf{x}\|) \leqslant c\}}\big)(g_\alpha(\|\mathbf{x}\|) - c)$, where $\tau \geqslant 1/2$, $c \geqslant 0$, $\alpha > 0$, and g_α is the Huber loss (Huber (1992)) defined by

$$g_\alpha(z) := \begin{cases} \frac{z^2}{2}, & |z| \leqslant \alpha, \\ \alpha(|z| - \frac{\alpha}{2}), & |z| > \alpha. \end{cases}$$

Then $\mathrm{Lip}(f) = \alpha\tau$ and $f(\mathbf{x}) = \tau(g_\alpha(\|\mathbf{x}\|) - c)$ on $\mathbb{R}^n \setminus \mathcal{K}$ with $\mathcal{K} := \{\mathbf{x} : g_\alpha(\|\mathbf{x}\|) \leqslant c\}$.

(iii) Log-exponential check loss: $f(\mathbf{x}) = \big(\tau - \mathbb{1}_{\{h(\|\mathbf{x}\|) \leqslant c\}}\big)(h(\|\mathbf{x}\|) - c)$, where $\tau \geqslant 1/2$, $c \geqslant 0$ and $h(z) := \log(1 + \exp(z))$. We have $\mathrm{Lip}(f) = \tau$ and $f(\mathbf{x}) = \tau(h(\|\mathbf{x}\|) - c)$ on $\mathbb{R}^n \setminus \mathcal{K}$ with $\mathcal{K} := \{\mathbf{x} : h(\|\mathbf{x}\|) \leqslant c\}$.

4 From Projection Equivalence to Regularization

In this section, we highlight a further payoff of projection equivalence: it pinpoints the pairs of risk measures ρ and loss functions f for which Wasserstein DRO admits an *exact regularization* representation. Such representations replace the worst-case problem by a nominal risk plus an explicit linear penalty, yielding a tractable and interpretable reformulation. Existing results of this type, however, are largely confined to affine losses f; see Wu et al. (2022). We show, at a general level, that the existence of an exact regularization representation is equivalent to projection equivalence. We say that (f, ρ) admits an *exact regularization* representation over the Wasserstein ball $\mathbb{B}_p(F_0, \varepsilon)$ if there exists a constant $c_{\text{reg}} > 0$ such that

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F\big(f(\boldsymbol{\xi})\big) = \rho^{F_0}\big(f(\boldsymbol{\zeta})\big) + c_{\text{reg}}\,\varepsilon, \tag{22}$$

where $\boldsymbol{\xi} \sim F$ and $\boldsymbol{\zeta} \sim F_0$. The constant c_{reg} generally depends on the underlying Wasserstein metric (e.g., its order and ground norm), as well as on ρ and f.

As a minimal prerequisite, we assume that ρ admits an exact regularization identity in one dimension. Specifically, for any baseline distribution G_0 on $\mathbb{R}$ and any $\varepsilon > 0$, there exists a constant $\bar{c} > 0$ such that

$$\sup_{G \in \mathbb{B}_p(G_0,\varepsilon)} \rho^G(\xi) = \rho^{G_0}(\zeta) + \bar{c}\,\varepsilon, \tag{23}$$

where $\xi \sim G$ and $\zeta \sim G_0$. Clearly, (23) is the one-dimensional special case of (22). Apply (23) with the reference distribution chosen as the pushforward $G_0 := f_{\#}F_0$ (i.e., the law of $f(\boldsymbol{\zeta})$ under $\boldsymbol{\zeta} \sim F_0$). This yields

$$\sup_{G \in \mathcal{C}_p(f \,|\, F_0, \tilde{c}\varepsilon)} \rho^G(X) = \rho^{F_0}\big(f(\boldsymbol{\zeta})\big) + c_{\text{reg}}\,\varepsilon,$$

where $c_{\text{reg}} = \tilde{c}\,\bar{c}$. Consequently, (22) holds if and only if

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F\big(f(\boldsymbol{\xi})\big) = \sup_{G \in \mathcal{C}_p(f \,|\, F_0, \tilde{c}\,\varepsilon)} \rho^G(X),$$

which is precisely projection equivalence.

Unlike the previous section, which established projection equivalence for broad classes of risk measures, we now restrict attention to risk measures satisfying (23). For each such family, we characterize the loss functions f for which projection equivalence holds. As noted in Section 3.3, the analysis bifurcates fundamentally between higher-order Wasserstein balls ($p > 1$) and the type-1 Wasserstein ball; we treat them separately in Sections 4.1 and 4.2, respectively.

4.1 Type-p Wasserstein Ball ($p > 1$)

L_p-norm risk

We begin with the higher-order L_p-risk functional

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \Big(\mathbb{E}^F\big[f^p(\boldsymbol{\xi})\big]\Big)^{1/p}, \tag{24}$$

where $f : \mathbb{R}^n \to \mathbb{R}_+$ and $\mathbb{B}_p(F_0, \varepsilon)$ denotes the type-p Wasserstein ball of radius ε centered at F_0.

A one-dimensional exact regularization identity is available for the absolute-value loss. In particular, Wu et al. (2022) shows that for any G_0 on $\mathbb{R}$ and any $\varepsilon > 0$,

$$\sup_{G \in \mathbb{B}_p(G_0,\varepsilon)} \Big(\mathbb{E}^G[|\xi|^p]\Big)^{1/p} = \Big(\mathbb{E}^{G_0}[|\zeta|^p]\Big)^{1/p} + \varepsilon, \tag{25}$$

where $\xi \sim G$ and $\zeta \sim G_0$. By setting $G_0 := f_{\#}F_0$, this in turn implies

$$\sup_{G \in \mathcal{C}_p(f \mid F_0, c_f \varepsilon)} \left(\mathbb{E}^G[|X|^p]\right)^{1/p} = \left(\mathbb{E}^{F_0}[f^p(\boldsymbol{\zeta})]\right)^{1/p} + c_f\,\varepsilon. \tag{26}$$

As noted earlier, an exact regularization counterpart of (24) follows once projection equivalence between (24) and (26) is established. For convex f, this condition can be characterized exactly.

Theorem 5. *Let $f : \mathbb{R}^n \to \mathbb{R}_+$ be convex and let $p \in (1, \infty)$. The following are equivalent:*

(i) There exists $c_f \geqslant 0$ such that for every $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$,

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \left(\mathbb{E}^F[f^p(\boldsymbol{\xi})]\right)^{1/p} = \sup_{G \in \mathcal{C}_p(f|F_0,c_f\varepsilon)} \left(\mathbb{E}^G[|X|^p]\right)^{1/p} = \left(\mathbb{E}^{F_0}[f^p(\boldsymbol{\zeta})]\right)^{1/p} + c_f\varepsilon. \tag{27}$$

(ii) There exist $\boldsymbol{\beta}_i \in \mathbb{R}^n$ with $\|\boldsymbol{\beta}_i\|_* = 1$ and $b_i \in \mathbb{R}$, $i \in I$, such that

$$f(\mathbf{x}) = \left(\max_{i \in I}\{c_f\,\boldsymbol{\beta}_i^\top \mathbf{x} + b_i\}\right)_+. \tag{28}$$

Two takeaways follow. First, once we specialize ρ to the L_p-risk functional, projection equivalence—and hence an exact regularization identity of the form (27)—holds for a class of losses that differs from that in Proposition 3. Second, the result is also an impossibility statement: an exact regularization counterpart is impossible for any convex loss function f that does not admit the representation (28). Notably, (28) subsumes, as special cases, the hinge-type losses and their variants studied in Theorem 4 and Corollary 1 of Wu et al. (2022).

Inf-form risk functionals

We next consider a broad class of inf-form risk functionals defined through an auxiliary scalar parameter. Specifically, for $p \in (1, \infty)$ we study

$$\mathcal{H}_p^F(X) = \inf_{t \in \mathbb{R}} \left\{ t + \left(\mathbb{E}^F\left[\ell^p(X,t)\right]\right)^{1/p} \right\}, \tag{29}$$

where $\ell : \mathbb{R} \times \mathbb{R} \to [0, \infty)$ is convex in its second argument t. To preclude degenerate cases and ensure that the infimum is attained at a finite value of t,[3] we impose the boundary–slope condition $\lim_{t \to -\infty} \partial_t \ell(z,t) < -1 < \lim_{t \to \infty} \partial_t \ell(z,t), \forall z \in \mathbb{R}$, with ∂_t interpreted as a subgradient

[3]These conditions exclude some trivial cases. For instance, if $\lim_{t \to -\infty} \partial_t \ell(z,t) \geqslant -1$, then $\ell(z,t)$ is increasing in t and the infimum of $t + (\mathbb{E}^F[\ell^p(X,t)])^{1/p}$ can be achieved at $t = -\infty$, rendering the problem uninformative. Analogous issues arise if $\lim_{t \to \infty} \partial_t \ell(z,t) \leqslant -1$. When ℓ is not everywhere differentiable in t, interpret ∂_t as any selection from the subdifferential and the limits as one-sided outer limits of subgradients.

selection when ℓ is not differentiable. This formulation encompasses many widely used risk measures, including higher-moment and related functionals.

For two canonical choices of ℓ, namely $\ell(z,t) = c(z-t)_+$ and $\ell(z,t) = c(z+|t|)_+$, an exact one-dimensional regularization identity is available; see Wu et al. (2022)

$$\sup_{G \in \mathbb{B}_p(G_0,\varepsilon)} \mathcal{H}_p^G(\xi) = \mathcal{H}_p^{G_0}(\zeta) + \varepsilon.$$

Combining this identity with the projection-equivalence mechanism developed earlier (Proposition 3), we obtain an immediate sufficient condition for exact regularization in higher dimensions, which extends the corresponding result in Wu et al. (2022).

Corollary 1. *For $p \in (1,\infty)$ and $c > 1$, let $\mathcal{H}_p$ be defined by (29) with loss ℓ. If $\ell(z,t) = c(z-t)_+$ or $\ell(z,t) = c(z+|t|)_+$, then for $\boldsymbol{\beta}_i \in \mathbb{R}^n$ with $\|\boldsymbol{\beta}_i\|_* = 1$ and $b_i \in \mathbb{R}$, $i \in I$, it holds that*

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \mathcal{H}_p^F\left(\max_{i\in I}\{\boldsymbol{\beta}_i^\top \boldsymbol{\xi} + b_i\}\right) = \mathcal{H}_p^{F_0}\left(\max_{i\in I}\{\boldsymbol{\beta}_i^\top \boldsymbol{\zeta} + b_i\}\right) + c\varepsilon. \tag{30}$$

As we show next, for the hinge-type loss $\ell(z,t) = c(z-t)_+$, which subsumes a number of standard higher-order risk measures as special cases, we can strengthen the preceding discussion by establishing an impossibility result.

Proposition 6. *For $p \in (1,\infty)$ and $c > 1$, let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $\mathcal{H}_p$ be defined by (29) with loss $\ell(z,t) = c(z-t)_+$. There exists $c_f \geqslant 0$ such that*

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \mathcal{H}_p^F\left(f(\boldsymbol{\xi})\right) = \sup_{G \in \mathcal{C}_p(f\,|\,F_0, c_f\varepsilon)} \mathcal{H}_p^G\left(Z\right) = \mathcal{H}_p^{F_0}\left(f(\boldsymbol{\zeta})\right) + c_f c\varepsilon \tag{31}$$

holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$ if and only if f satisfies (12).

Distortion risk functionals

We now turn to distortion risk functionals, a classical and widely used generalization of the expectation (Yaari (1987), Schmeidler (1989)). A risk measure ρ_h is called a *distortion risk measure* if

$$\rho_h^F(Z) = \int_0^1 \mathrm{VaR}_u^F(Z)\,\mathrm{d}h(u),$$

where $h : [0,1] \to [0,1]$ is increasing with $h(0) = 0$ and $h(1) = 1$. Throughout, we focus on convex distortions h, for which ρ_h is coherent, i.e., $\rho_h \in \mathcal{R}_{\mathrm{coh}}$. We further define

$$\|h'\|_q := \left(\int_0^1 (h'(u))^q \, \mathrm{d}u \right)^{1/q},$$

where h' denotes the left derivative of h.

A one-dimensional exact regularization identity is available for ρ_h (see, e.g., Wu et al. (2022)):

$$\sup_{G \in \mathbb{B}_p(G_0, \varepsilon)} \rho_h^G(\xi) = \rho_h^{G_0}(\zeta) + \varepsilon \, \|h'\|_q, \tag{32}$$

where $\xi \sim G$ and $\zeta \sim G_0$. Combining the projection-equivalence result in Theorem 3 with the identity above yields an immediate sufficient condition for an exact regularization reformulation. Less obviously, the result below establishes sharpness: even when attention is restricted to distortion risk measures, the structural condition in Theorem 6 remains necessary. Consequently, an exact regularization counterpart is impossible outside that class.

Theorem 6. *For $p > 1$, let h be a convex distortion function satisfying $\|h'\|_q \in \mathbb{R}$, and $f : \mathbb{R}^n \to \mathbb{R}$ be a function. There exists $c_f \geqslant 0$ such that*

$$\sup_{F \in \mathbb{B}_p(F_0, \varepsilon)} \rho_h^F(f(\boldsymbol{\xi})) = \sup_{G \in \mathcal{C}_p(f | F_0, c_f \varepsilon)} \rho_h^G(X) = \rho_h^{F_0}(f(\boldsymbol{\zeta})) + c_f \varepsilon \|h'\|_q \tag{33}$$

holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$ if and only if f is given by (11).

4.2 Type-1 Wasserstein Ball

The type-1 case is qualitatively different. As noted in Section 3.3, projection equivalence can hold for a substantially broader class of losses f when $p = 1$ than when $p > 1$. We show that the same relaxation carries over to the risk measures considered above: in the type-1 setting, exact regularization holds under weaker conditions on f, leading to a strictly larger admissible class.

Inf-form risk functionals

We first revisit the inf-form functionals in (29) in the type-1 setting. Specializing to $p = 1$, the functional $\mathcal{H}_1$ defined in (29) admits a convenient expectation form. For a given loss ℓ, define $\ell_1(z, t) := t + \ell(z, t)$. Then

$$\mathcal{H}_1^F(X) = \inf_{t \in \mathbb{R}} \mathbb{E}^F\big[\ell_1(X, t)\big]. \tag{34}$$

We assume that $\ell_1 : \mathbb{R}^2 \to \mathbb{R}$ is convex in its second argument t and satisfies the boundary–slope condition $\lim_{t\to-\infty} \partial_t \ell_1(z,t) < 0 < \lim_{t\to\infty} \partial_t \ell_1(z,t), \forall z \in \mathbb{R}$, which ensures that the infimum in (34) is non-degenerate and attained at a finite value of t.

Even in one dimension, exact regularization identities for the worst-case problems in (34) are available only for a few special losses ℓ_1. For a general ℓ_1, it remains largely open to characterize which loss functions f admit an exact regularization counterpart. We do not pursue a complete characterization here; instead, we provide a broad sufficient condition.

Proposition 7. *Let $f : \mathbb{R}^n \to \mathbb{R}$ satisfy the condition of Theorem 4(i), and let $\mathcal{H}_1$ be given in (34) with loss function $\ell_1 : \mathbb{R}^2 \to \mathbb{R}$ satisfying $\mathrm{Lip}(\ell_1(\cdot, t)) = b \in \mathbb{R}_+$ for all $t \in \mathbb{R}$ and for each $t \in \mathbb{R}$ there exists $z_0(t)$ such that*

$$\lim_{m\to\infty} \frac{\ell_1(z_0(t) + m, t) - \ell_1(z_0(t), t)}{m} = b. \tag{35}$$

It holds that for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$,

$$\sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \mathcal{H}_1^F(f(\boldsymbol{\xi})) = \mathcal{H}_1^{F_0}(f(\boldsymbol{\zeta})) + b\,\mathrm{Lip}(f)\varepsilon. \tag{36}$$

Note that when ℓ_1 is such that the induced risk functional $\mathcal{H}_1$ is coherent, the regularization results above follow directly by combining the one-dimensional identity with the projection-equivalence result in Theorem 4. Proposition 7, however, applies more broadly: depending on ℓ_1, the functional $\mathcal{H}_1$ in (34) need not be coherent and may not even be monotone.

Condition (35) is mild. In particular, if for every $t \in \mathbb{R}$ the map $z \mapsto \ell_1(z,t)$ is convex and satisfies $\mathrm{Lip}(\ell_1(\cdot,t)) = b$, then (35) holds automatically. Many standard loss functions satisfy this requirement, along with the assumptions of Proposition 7.

Example 5. The following are concrete examples of functions ℓ_1 that satisfy the assumptions in Proposition 7.

(i) Quantile loss (Koenker and Bassett (1978)): $\ell_1(z,t) = \alpha(z-t)_+ + (1-\alpha)(z-t)_-$, $\alpha \in [1/2, 1)$, which satisfies $\mathrm{Lip}(\ell_1(\cdot,t)) = \alpha$ for every t and (35) holds with $b = \alpha$. In this case, $\mathcal{H}_1^F(X)$ reduces to the expected check loss.

(ii) Huber loss (Huber (1992)): For $\alpha > 0$,

$$\ell_1(z,t) = \begin{cases} \frac{(z-t)^2}{2}, & |z-t| \leqslant \alpha, \\ \alpha(|z-t| - \frac{\alpha}{2}), & |z-t| > \alpha, \end{cases}$$

which satisfies $\mathrm{Lip}(\ell_1(\cdot,t)) = \alpha$ for every t and (35) holds with $b = \alpha$.

(iii) Pseudo-Huber loss: For $\alpha > 0$, $\ell_1(z,t) = \alpha^2\left(\sqrt{1+((z-t)/\alpha)^2} - 1\right)$, which satisfies $\mathrm{Lip}(\ell_1(\cdot,t)) = \alpha$ for every t and (35) holds with $b = \alpha$.

Distortion risk functionals

We next revisit distortion risk measures in the type-1 setting. In contrast to the higher-order case $p > 1$, where exact regularization typically forces a rigid structure on the loss f, the type-1 geometry allows exact regularization to hold for a substantially broader class of losses. In particular, worst-case distortion risks over $\mathbb{B}_1(F_0,\varepsilon)$ admit regularization counterparts under markedly weaker requirements on f than those identified in Section 4.1. The following result provides a general sufficient condition, obtained by combining the one-dimensional regularization identity for distortion risk measures (see, e.g., Kuhn et al. (2025), Wu et al. (2022)) with the projection-equivalence result in Theorem 4.

Proposition 8. *Let h be an increasing and convex distortion function satisfying $\|h'\|_\infty \in (0,\infty)$, and let $f : \mathbb{R}^n \to \mathbb{R}$ satisfy the conditions of Theorem 4 (i). We have*

$$\sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \rho_h^F\left(f\left(\boldsymbol{\xi}\right)\right) = \rho_h^{F_0}\left(f\left(\boldsymbol{\zeta}\right)\right) + \mathrm{Lip}(f)\varepsilon\left\|h'\right\|_\infty.$$

holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$.

Expectile risk functionals

Before moving on, we pause to highlight that the type-1 setting supports more than the exact regularization identities derived for the specific functionals above. Once projection equivalence is available, it can also serve as a general device for obtaining other "regularization-like" reformulations for coherent risk measures. To illustrate, we consider the expectile (Newey and Powell (1987)). For a risk X with distribution F, the α-expectile $\mathrm{ex}_\alpha^F(X)$ is defined as the unique solution x to

$$\alpha\,\mathbb{E}^F[(X-x)_+] = (1-\alpha)\,\mathbb{E}^F[(X-x)_-], \tag{37}$$

and it is coherent when $\alpha \geqslant \frac{1}{2}$ (Bellini et al. (2014)). By Theorem 4, if f satisfies the conditions of Theorem 4(i), then

$$\sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \mathrm{ex}_\alpha^F\big(f(\boldsymbol{\xi})\big) = \sup_{G\in\mathcal{C}_1(f\,|\,F_0,\,\mathrm{Lip}(f)\varepsilon)} \mathrm{ex}_\alpha^G(X).$$

The next proposition shows that the worst-case expectile can be characterized equivalently as the solution to a regularized version of (37).

Proposition 9. *For $\alpha \in [1/2, 1]$, $F_0 \in \mathcal{M}(\mathbb{R}^n)$, $\varepsilon > 0$, and a function f satisfying the conditions in Theorem 4(i), we have $\sup_{F \in \mathbb{B}_1(F_0,\varepsilon)} \mathrm{ex}_\alpha^F(f(\boldsymbol{\xi}))$ is the unique solution to*

$$\mathbb{E}^{F_0}[\alpha(f(\boldsymbol{\zeta}) - x)_+ - (1-\alpha)(f(\boldsymbol{\zeta}) - x)_-] + \alpha \mathrm{Lip}(f)\varepsilon = 0.$$

5 Classification

As a natural extension of the Wasserstein DRO problem (1), we now turn to a setup motivated by classification in machine learning. Here, the random vector consists of a class label and a feature vector, $\boldsymbol{\xi} = (Y, \mathbf{X}) \in \Xi$, where $\Xi := \{-1, 1\} \times \mathbb{R}^n \subseteq \mathbb{R}^{n+1}$, with Y denoting a binary label and $\mathbf{X}$ the associated features. The classification task is to select a decision function $f : \mathbb{R}^n \to \mathbb{R}$, interpreted as a classifier, from a class $\mathcal{A}$ to predict the sign of Y given $\mathbf{X}$.

To capture distributional robustness, we equip Ξ with the type-p Wasserstein metric

$$\overline{\mathrm{W}}_p(F_1, F_2) := \inf_{\pi \in \overline{\Pi}(F_1,F_2)} \left(\mathbb{E}^\pi[d(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)^p]\right)^{1/p},$$

where $\overline{\Pi}(F_1, F_2)$ is the set of all distributions on Ξ with marginals F_1 and F_2 supported on Ξ, the distance between $\boldsymbol{\xi}_1 = (Y_1, \mathbf{X}_1)$ and $\boldsymbol{\xi}_2 = (Y_2, \mathbf{X}_2)$ is defined via the additively separable form

$$d(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \|\mathbf{X}_1 - \mathbf{X}_2\| + \Theta(Y_1 - Y_2), \tag{38}$$

with $\|\cdot\|$ a norm on $\mathbb{R}^n$. The penalty function $\Theta : \mathbb{R} \to \{0, \infty\}$ is specified by $\Theta(0) = 0$ and $\Theta(s) = \infty$ for $s \neq 0$. Hence, the metric prohibits perturbations in the label Y while allowing adversarial shifts in the feature space $\mathbf{X}$.

For a nominal distribution $F_0 \in \mathcal{M}_p(\Xi)$ and robustness radius $\varepsilon > 0$, the distributionally robust classification problem is given by

$$\inf_{f \in \mathcal{A}} \sup_{F \in \overline{\mathbb{B}}_p(F_0,\varepsilon)} \rho^F\big(Y \cdot f(\mathbf{X})\big),$$

where $\overline{\mathbb{B}}_p(F_0, \varepsilon) := \{F \in \mathcal{M}(\Xi) : \overline{\mathrm{W}}_p(F, F_0) \leqslant \varepsilon\}$, and ρ is a risk functional applied to the *margin* $Z := Y \cdot f(\mathbf{X})$.

To study projection equivalence in this setting, let $(Y_0, \mathbf{X}_0) \sim F_0$ and denote by G_0 the distribution of the baseline margin $Y_0 \cdot f(\mathbf{X}_0)$. We introduce the one-dimensional Wasserstein ball

$$\overline{\mathcal{C}}_p(f|F_0,\varepsilon) := \mathbb{B}_p(G_0,\varepsilon) \subseteq \mathcal{M}(\mathbb{R}).$$

A classifier f admits *classification projection equivalence* if

$$\sup_{F\in\overline{\mathbb{B}}_p(F_0,\varepsilon)} \rho^F(Y\cdot f(\mathbf{X})) = \sup_{G\in\overline{\mathcal{C}}_p(f|F_0,c_f\varepsilon)} \rho^G(Z)$$

for some constant $c_f \geqslant 0$.

As shown below, in contrast to projection equivalence in Section 3, classification projection equivalence holds only for classifiers f that strictly satisfy set-level equivalence. This requirement remains even when ρ is restricted to the monotone class $\mathcal{R}_{\mathrm{mon}}$, underscoring that exact reduction in classification demands stronger conditions than in the typical Wasserstein DRO framework (1).

Proposition 10. *For $p \geqslant 1$ and $\alpha \in [0,1)$, let $f:\mathbb{R}^n \to \mathbb{R}$ be a function. The following statements are equivalent.*

(i) There exists $c_f \geqslant 0$ such that

$$\sup_{F\in\overline{\mathbb{B}}_p(F_0,\varepsilon)} \rho^F\left(Y\cdot f(\mathbf{X})\right) = \sup_{G\in\overline{\mathcal{C}}_p(f|F_0,c_f\varepsilon)} \rho^G(X) \tag{39}$$

holds for any $\rho \in \mathcal{R}_{\mathrm{mon}}$, $F_0 \in \mathcal{M}(\Xi)$, and $\varepsilon > 0$.

(ii) There exists $c_f \geqslant 0$ such that for any $F_0 \in \mathcal{M}(\Xi)$ and $\varepsilon > 0$, it holds that

$$\{F_{Y\cdot f(\mathbf{X})}:\ F_{(Y,\mathbf{X})} \in \overline{\mathbb{B}}_p(F_0,\varepsilon)\} = \overline{\mathcal{C}}_p(f|F_0,c_f\varepsilon). \tag{40}$$

(iii) The function f is Lipschitz continuous, and satisfies (8).

Moreover, this necessity remains even under specialized risk measures; in particular, the characterization (8) is still required when the risk functional is Value-at-Risk.

Proposition 11. *For $\alpha \in [0,1)$ and $p \geqslant 1$, let $f:\mathbb{R}^n \to \mathbb{R}$ be a function. There exists $c_f \geqslant 0$ such that*

$$\sup_{F\in\overline{\mathbb{B}}_p(F_0,\varepsilon)} \mathrm{VaR}^F_\alpha\left(Y\cdot f(\mathbf{X})\right) = \sup_{G\in\overline{\mathcal{C}}_p(f|F_0,c_f\varepsilon)} \mathrm{VaR}^G_\alpha(X) \tag{41}$$

holds for any $F_0 \in \mathcal{M}(\Xi)$ and $\varepsilon > 0$ if and only if f satisfies (8).

Recall that when the norm $\|\cdot\| = \|\cdot\|_a$ is the ℓ_a-norm for some $a \in [1,\infty)$, Propositions 1 and 2 identify explicit loss-function forms under which projection equivalence holds. Combining these loss-function forms with Proposition 10 and the one-dimensional regularization representation for convex distortion risk measures, i.e. (32), yields the following regularization reformulation for distributionally robust classification under distortion risk measures (see Wu et al. (2022) for examples of such classification formulations), linking robust classification directly to a familiar paradigm in machine learning.

Corollary 2. *For $p \geqslant 1$, $F_0 \in \mathcal{M}(\Xi)$ with $(Y_0, \mathbf{X}_0) \sim F_0$, let h be an increasing and convex distortion function satisfying $\|h'\|_q \in (0,\infty)$. We have the following statements.*

(i) If $f(\mathbf{x}) = \boldsymbol{\beta}^\top \mathbf{x} + b$ for some $\boldsymbol{\beta} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, we have

$$\sup_{F \in \overline{\mathbb{B}}_p(F_0,\varepsilon)} \rho_h^F(Y \cdot f(\mathbf{X})) = \rho_h^{F_0}(Y_0 \cdot f(\mathbf{X}_0)) + \|\boldsymbol{\beta}\|_* \varepsilon \|h'\|_q.$$

(ii) If $\|\cdot\| = \|\cdot\|_1$ is the ℓ_1-norm and f is given by Proposition 1, then we have

$$\sup_{F \in \overline{\mathbb{B}}_p(F_0,\varepsilon)} \rho_h^F(Y \cdot f(\mathbf{X})) = \rho_h^{F_0}(Y_0 \cdot f(\mathbf{X}_0)) + c\varepsilon \|h'\|_q.$$

We note that while case (i) can also be obtained through a direct analysis of linear classifiers, as shown in Wu et al. (2022), case (ii) emerges only within the more general framework of Proposition 10.

6 Conclusion

In this work, we provide the first complete characterization of *projection equivalence* in Wasserstein distributionally robust optimization. Our central finding reveals that this powerful high-dimensional reduction is not confined to the restrictive case of set-level equivalence but extends to a much broader class of loss functions. By systematically navigating a hierarchy of risk functionals, we establish a sharp boundary delineating precisely when a high-dimensional worst-case risk evaluation simplifies to its one-dimensional counterpart. This foundational result, in turn, enables us to derive necessary and sufficient conditions for Wasserstein DRO problems to admit exact regularization reformulations, unifying two central paradigms in optimization and machine learning. Ultimately, our analysis delivers new classes of tractable models and establishes the fundamental limits of such reductions, clarifying the structural properties that govern computational feasibility in distributionally robust optimization.

A Appendix: Proofs of the Main results

A.1 Proofs for Section 3

To prove Theorem 1, we need the following lemma.

Lemma A1. *For any $p \geqslant 1$, $\varepsilon > 0$, $\alpha \in [0,1)$ and $f : \mathbb{R}^n \to \mathbb{R}$, we have*

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p]\leqslant\varepsilon^p} \mathrm{VaR}_\alpha\big(f(\boldsymbol{\xi})\big) = \sup_{\|\mathbf{z}-\mathbf{x}\|\leqslant\varepsilon/(1-\alpha)^{1/p}} f(\mathbf{z}), \qquad \forall\, \mathbf{x} \in \mathbb{R}^n. \tag{A1}$$

Proof. Denote by $\varepsilon_0 = \varepsilon/(1-\alpha)^{1/p}$. Note that for any $\mathbf{z} \in \mathbb{R}^n$ with $\|\mathbf{z}-\mathbf{x}\| \leqslant \varepsilon_0$, we can define $\boldsymbol{\xi} = \mathbf{x}$ if $f(\mathbf{z}) \leqslant f(\mathbf{x})$ and $\boldsymbol{\xi} \sim \alpha\delta_{\mathbf{x}} + (1-\alpha)\delta_{\mathbf{z}}$ if $f(\mathbf{z}) > f(\mathbf{x})$. In both cases, we have

$$\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p] \leqslant (1-\alpha)\|\mathbf{z}-\mathbf{x}\|^p \leqslant \varepsilon^p \text{ and } \mathrm{VaR}_\alpha(f(\boldsymbol{\xi})) = \max\{f(\mathbf{z}), f(\mathbf{x})\}.$$

Thus,

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p]\leqslant\varepsilon^p} \mathrm{VaR}_\alpha(f(\boldsymbol{\xi})) \geqslant \sup_{\|\mathbf{z}-\mathbf{x}\|\leqslant\varepsilon_0} f(\mathbf{z}).$$

We next prove the converse inequality. For any $\boldsymbol{\xi}$ with $\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p] \leqslant \varepsilon^p$, note that if $\mathbb{P}\left(\|\boldsymbol{\xi}-\mathbf{x}\| > \varepsilon_0\right) > 0$, then

$$\varepsilon_0^p\mathbb{P}\left(\|\boldsymbol{\xi}-\mathbf{x}\| > \varepsilon_0\right) < \mathbb{E}\left[\|\boldsymbol{\xi}-\mathbf{x}\|^p\mathbb{1}_{\{\|\boldsymbol{\xi}-\mathbf{x}\|>\varepsilon_0\}}\right] \leqslant \mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p] \leqslant \varepsilon^p.$$

It follows that $\mathbb{P}(\|\boldsymbol{\xi}-\mathbf{x}\| > \varepsilon_0) < 1-\alpha$, that is, $\mathbb{P}(\|\boldsymbol{\xi}-\mathbf{x}\| \leqslant \varepsilon_0) > \alpha$. Thus

$$\mathbb{P}\left(f(\boldsymbol{\xi}) \leqslant \sup_{\|\mathbf{z}-\mathbf{x}\|\leqslant\varepsilon_0} f(\mathbf{z})\right) \geqslant \mathbb{P}(\|\boldsymbol{\xi}-\mathbf{x}\| \leqslant \varepsilon_0) > \alpha.$$

By the definition of VaR, it follows that $\mathrm{VaR}_\alpha(f(\boldsymbol{\xi})) \leqslant \sup_{\|\mathbf{z}-\mathbf{x}\|\leqslant\varepsilon_0} f(\mathbf{z})$ for any $\boldsymbol{\xi}$ with $\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p] \leqslant \varepsilon^p$. Therefore, the converse direction holds and we complete the proof. □

Proof of Theorem 1. Note that the implication $(ii) \Rightarrow (i)$ is trivial. We only give the proof of $(i) \Rightarrow (iii)$ and $(iii) \Rightarrow (ii)$.

For $(i) \Rightarrow (iii)$, we first consider the case $c_f = 0$. Choose $\rho = \mathrm{VaR}_\alpha$ for some $\alpha \in [0,1)$ and take

$F_0 = \delta_{\mathbf{x}}$ for $\mathbf{x} \in \mathbb{R}^n$. In this case, (6) reduces to

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p]\leqslant\varepsilon^p} \mathrm{VaR}_\alpha\left(f(\boldsymbol{\xi})\right) = \sup_{\mathbb{E}[|X-f(\mathbf{x})|^p]\leqslant 0} \mathrm{VaR}_\alpha\left(X\right) = f(\mathbf{x}). \tag{A2}$$

This, together with (A1) in Lemma A1, implies

$$\sup_{\|\mathbf{z}-\mathbf{x}\|\leqslant\varepsilon/(1-\alpha)^{1/p}} f(\mathbf{z}) = f(\mathbf{x}), \quad \forall\, \varepsilon > 0.$$

By $\mathbf{x}$ is arbitrary, we have f is a constant function, which completes the proof for $c_f = 0$. We next consider the case where $c_f > 0$. Without loss of generality, set $c_f = 1$. For $\rho = \mathrm{VaR}_\alpha$ with $\alpha \in [0, 1)$, we first show that (6) implies that

$$\sup_{\|\mathbf{y}\|\leqslant\varepsilon} f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) = \varepsilon, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\ \varepsilon > 0. \tag{A3}$$

To that end, take $F_0 = \delta_{\mathbf{x}}$ for $\mathbf{x} \in \mathbb{R}^n$. In this case, (6) reduces to

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{x}\|^p]\leqslant\varepsilon^p} \mathrm{VaR}_\alpha(f(\boldsymbol{\xi})) = \sup_{\mathbb{E}[|Z-f(\mathbf{x})|^p]\leqslant\varepsilon^p} \mathrm{VaR}_\alpha(Z). \tag{A4}$$

By Lemma A1, we have (A1) and

$$\sup_{\mathbb{E}[|Z-f(\mathbf{x})|^p]\leqslant\varepsilon^p} \mathrm{VaR}_\alpha(Z) = \sup_{|z-f(\mathbf{x})|\leqslant\frac{\varepsilon}{(1-\alpha)^{1/p}}} z = f(\mathbf{x}) + \frac{\varepsilon}{(1-\alpha)^{1/p}}. \tag{A5}$$

Combining (A1), (A5) and (A4) yields

$$\sup_{\|\mathbf{z}-\mathbf{x}\|\leqslant\frac{\varepsilon}{(1-\alpha)^{1/p}}} f(\mathbf{z}) = f(\mathbf{x}) + \frac{\varepsilon}{(1-\alpha)^{1/p}}.$$

For $\alpha \in [0, 1)$, since $\varepsilon > 0$ is arbitrary, it follows that for any $\varepsilon > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\sup_{\|\mathbf{z}-\mathbf{x}\|\leqslant\varepsilon} f(\mathbf{z}) = f(\mathbf{x}) + \varepsilon,$$

that is, (A3) holds. This identity immediately implies that

$$|f(\mathbf{x}) - f(\mathbf{y})| \leqslant \|\mathbf{x} - \mathbf{y}\|, \quad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

and thus f is Lipschitz continuous with Lipschitz constant $\mathrm{Lip}(f) = 1$. Moreover, since (6) holds for any risk measure, by taking $\rho = -\mathrm{VaR}_\alpha$ and repeating a similar argument, we obtain that for all $\varepsilon > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) - \inf_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x} + \mathbf{y}) = \varepsilon, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\ \varepsilon > 0,$$

and thus (8) follows. This completes the proof of $(i) \Rightarrow (iii)$.

For $(iii) \Rightarrow (ii)$, first note that the case $\mathrm{Lip}(f) = 0$ is trivial as (7) holds with $c_f = 0$. We next consider the case $\mathrm{Lip}(f) > 0$. Assume without loss of generality that $\mathrm{Lip}(f) = 1$. We aim to show that (7) holds for $c_f = 1$. Since f is 1-Lipschitz continuous, we have for any $\boldsymbol{\xi}, \boldsymbol{\zeta}$ with $\mathbb{E}[\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p] \leqslant \varepsilon^p$, it holds that

$$\mathbb{E}\left[|f(\boldsymbol{\xi}) - f(\boldsymbol{\zeta})|^p\right] \leqslant \mathbb{E}[\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p] \leqslant \varepsilon^p,$$

which implies $f_\# \mathbb{B}_p(F_0, \varepsilon) \subseteq \mathcal{C}_p(f|F_0, \varepsilon)$. To prove the reverse inclusion, it suffices to show that for any $G \in \mathcal{C}_p(f|F_0, \varepsilon)$ and $Z \sim G$ with $\mathbb{E}[|Z - f(\boldsymbol{\zeta})|^p] \leqslant \varepsilon^p$, there exists a random vector $\boldsymbol{\xi}$ with distribution $F_{\boldsymbol{\xi}} \in \mathbb{B}_p(F_0, \varepsilon)$ such that $f(\boldsymbol{\xi}) = Z$ almost surely. To this end, denote by $T := Z - f(\boldsymbol{\zeta})$. Then, we have $\mathbb{E}\left[|T|^p\right] \leqslant \varepsilon^p$. By the measurable selection theorem (See Theorem 3.5 in Rieder (1978); see also Lemma EC.12 in Wu et al. (2022)), there exist measurable mappings $\mathbf{V}_1$ and $\mathbf{V}_2$ such that

$$\mathbf{V}_1(\omega) \in \operatorname*{arg\,max}_{\|\mathbf{y}\| \leqslant |T(\omega)|} f\left(\boldsymbol{\zeta}(\omega) + \mathbf{y}\right) \quad \text{and} \quad \mathbf{V}_2(\omega) \in \operatorname*{arg\,min}_{\|\mathbf{y}\| \leqslant |T(\omega)|} f\left(\boldsymbol{\zeta}(\omega) + \mathbf{y}\right),\ \omega \in \Omega.$$

Then, denote by $A_+ := \{\omega : T(\omega) \geqslant 0\}$ and $A_- := \{\omega : T(\omega) < 0\}$ and define

$$\boldsymbol{\xi}(\omega) = (\boldsymbol{\zeta}(\omega) + \mathbf{V}_1(\omega)) \mathbb{1}_{A_+}(\omega) + (\boldsymbol{\zeta}(\omega) + \mathbf{V}_2(\omega)) \mathbb{1}_{A_-}(\omega), \quad \omega \in \Omega,$$

which is measurable. As $\Omega = A_+ \cup A_-$, we have

$$\|\boldsymbol{\xi}(\omega) - \boldsymbol{\zeta}(\omega)\| \leqslant \max\left\{\|\mathbf{V}_1(\omega)\|, \|\mathbf{V}_2(\omega)\|\right\} \leqslant |T(\omega)|, \quad \omega \in \Omega,$$

and thus, $\mathbb{E}[\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p] \leqslant \mathbb{E}[|T|^p] \leqslant \varepsilon^p$. This implies $F_{\boldsymbol{\xi}} \in \mathbb{B}_p(F_0, \varepsilon)$. Moreover, we have for $\omega \in A_+$

$$\begin{aligned} f\left(\boldsymbol{\xi}(\omega)\right) &= f(\boldsymbol{\zeta}(\omega) + \mathbf{V}_1(\omega)) \\ &= \max_{\|\mathbf{y}\| \leqslant |T(\omega)|} f(\mathbf{y} + \boldsymbol{\zeta}(\omega)) = f\left(\boldsymbol{\zeta}(\omega)\right) + T(\omega) = Z(\omega), \end{aligned}$$

where the second equality follows from the definition of $\mathbf{V}_1$, the third equality follows from (8), and the last equality follows from the definition of T, that is, $T = Z - f\left(\boldsymbol{\zeta}\right)$. Similarly, one can verify that for $\omega \in A_-$,

$$f(\boldsymbol{\xi}(\omega)) = f(\boldsymbol{\zeta}(\omega) + \mathbf{V}_2(\omega)) = f(\boldsymbol{\zeta}(\omega)) - (-T(\omega)) = Z(\omega).$$

Therefore, $\mathcal{C}_p(f|F_0, \varepsilon) \subseteq f_{\#}\mathbb{B}_p(F_0, \varepsilon)$, and (7) follows, completing the proof.

□

To prove Proposition 1, we need the following two lemmas, and the first lemma will be used repeatedly throughout the appendix.

Lemma A2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz continuous function. We have the following statements hold.*

(i) If f satisfies

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) = \mathrm{Lip}(f)\varepsilon, \ \forall \, \mathbf{x}, \, \varepsilon \in \mathbb{R}_+, \tag{A6}$$

then for each $\mathbf{x} \in \mathbb{R}^n$, there exists $\boldsymbol{\beta}_{\mathbf{x}}$ such that $\|\boldsymbol{\beta}_{\mathbf{x}}\| = 1$ and

$$f(\mathbf{x} + \varepsilon \boldsymbol{\beta}_{\mathbf{x}}) - f(\mathbf{x}) = \mathrm{Lip}(f)\varepsilon, \qquad \forall \, \varepsilon > 0. \tag{A7}$$

(ii) If f satisfies

$$f(\mathbf{x}) - \inf_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x} + \mathbf{y}) = \mathrm{Lip}(f)\varepsilon, \ \forall \, \mathbf{x}, \, \varepsilon \in \mathbb{R}_+,$$

then for each $\mathbf{x} \in \mathbb{R}^n$, there exists $\boldsymbol{\eta}_{\mathbf{x}}$ such that $\|\boldsymbol{\eta}_{\mathbf{x}}\| = 1$ and

$$f(\mathbf{x}) - f(\mathbf{x} + \varepsilon \boldsymbol{\eta}_{\mathbf{x}}) = \mathrm{Lip}(f)\varepsilon, \qquad \forall \, \varepsilon > 0.$$

Proof. (i) Without loss of generality, assume that $\mathrm{Lip}(f) = 1$ and fix $\mathbf{x} \in \mathbb{R}^n$. By (A6) and the Lipschitz continuouity, for each $\varepsilon > 0$, there exists $\boldsymbol{\beta}_{\mathbf{x},\varepsilon}$ such that $\|\boldsymbol{\beta}_{\mathbf{x},\varepsilon}\| = 1$ and

$$f(\mathbf{x} + \varepsilon \boldsymbol{\beta}_{\mathbf{x},\varepsilon}) - f(\mathbf{x}) = \varepsilon \|\boldsymbol{\beta}_{\mathbf{x},\varepsilon}\| = \varepsilon. \tag{A8}$$

We assert that $\boldsymbol{\beta}_{\mathbf{x},\varepsilon}$ can be chosen independently of ε. To show this, first note that for fixed $\varepsilon > 0$

and the chosen $\boldsymbol{\beta}_{\mathbf{x},\varepsilon}$, (A8) implies that

$$f(\mathbf{x}+\varepsilon'\boldsymbol{\beta}_{\mathbf{x},\varepsilon})-f(\mathbf{x})=\varepsilon'\|\boldsymbol{\beta}_{\mathbf{x},\varepsilon}\|=\varepsilon',\quad \forall\,\varepsilon'\in[0,\varepsilon]. \tag{A9}$$

This is due to that if (A9) does not hold for some $\varepsilon'\in[0,\varepsilon]$, then we have $f(\mathbf{x}+\varepsilon'\boldsymbol{\beta}_{\mathbf{x},\varepsilon})-f(\mathbf{x})<\varepsilon'$ by the 1-Lipschitz continuity of f, and thus,

$$\begin{aligned}f(\mathbf{x}+\varepsilon\boldsymbol{\beta}_{\mathbf{x},\varepsilon})-f(\mathbf{x})&=f(\mathbf{x}+\varepsilon\boldsymbol{\beta}_{\mathbf{x},\varepsilon})-f(\mathbf{x}+\varepsilon'\boldsymbol{\beta}_{\mathbf{x},\varepsilon})+f(\mathbf{x}+\varepsilon'\boldsymbol{\beta}_{\mathbf{x},\varepsilon})-f(\mathbf{x})\\&\leqslant\varepsilon-\varepsilon'+f(\mathbf{x}+\varepsilon'\boldsymbol{\beta}_{\mathbf{x},\varepsilon})-f(\mathbf{x})<\varepsilon,\end{aligned}$$

where the first inequality follows from that f is 1-Lipschitz continuous. This yields a contradiction to (A8). Therefore, by (A9), we have for each $\varepsilon_n\geqslant n$, $n\in\mathbb{N}$, there exists $\boldsymbol{\beta}_n\in\mathbb{R}^n$ such that $\|\boldsymbol{\beta}_n\|=1$ and

$$f(\mathbf{x}+\varepsilon\boldsymbol{\beta}_n)-f(\mathbf{x})=\varepsilon,\quad\forall\,\varepsilon\in[0,\varepsilon_n].$$

By the Bolzano–Weierstrass theorem, for $\|\boldsymbol{\beta}_n\|\leqslant 1$, $n\in\mathbb{N}$, there exists $\boldsymbol{\beta}_{\mathbf{x}}$ such that $\|\boldsymbol{\beta}_n-\boldsymbol{\beta}_{\mathbf{x}}\|\to 0$ as $n\to\infty$. We thus have $\|\boldsymbol{\beta}_{\mathbf{x}}\|=1$ and by the 1-Lipschitz continuouity of f again,

$$f(\mathbf{x}+\varepsilon\boldsymbol{\beta}_{\mathbf{x}})-f(\mathbf{x})=\varepsilon,\quad\forall\,\varepsilon\geqslant 0.$$

Thus, this completes the proof of (i).

(ii) The proof is similar to that of (i), and thus, is omitted. □

Lemma A3. *If $\|\cdot\|=\|\cdot\|_1$ is the ℓ_1-norm, then f satisfies (8) if and only if for any $\mathbf{x}\in\mathbb{R}^n$, there exist $\tilde{\mathbf{e}}_i\in\{\pm\mathbf{e}_i\}$ and $\tilde{\mathbf{e}}_j\in\{\pm\mathbf{e}_j\}$ for some $i,j\in[n]$ such that*

$$f(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i)-f(\mathbf{x})=f(\mathbf{x})-f(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_j)=\mathrm{Lip}(f)\varepsilon,\quad\forall\,\varepsilon>0; \tag{A10}$$

Proof. Without loss of generality, assume that $\mathrm{Lip}(f)=1$. If $\|\cdot\|=\|\cdot\|_1$, then by Lemma A2, (A7) holds. That is, for any $\mathbf{x}\in\mathbb{R}^n$, there exists $\boldsymbol{\beta}_{\mathbf{x}}$ with $\|\boldsymbol{\beta}_{\mathbf{x}}\|_1=1$ such that

$$f(\mathbf{x}+\varepsilon\boldsymbol{\beta}_{\mathbf{x}})-f(\mathbf{x})=\varepsilon=\varepsilon\sum_{k=1}^{n}|\beta_{\mathbf{x}k}|,\quad\forall\,\varepsilon>0. \tag{A11}$$

Denote by $\mathbf{x}_0:=\mathbf{x}$ and $\mathbf{x}_k:=\mathbf{x}_{k-1}+\varepsilon\beta_{\mathbf{x}k}\mathbf{e}_k$, $k\in[n]$, where $\beta_{\mathbf{x}k}$ denotes the k-th component of $\boldsymbol{\beta}_{\mathbf{x}}$.

By Lipschitz continuity,

$$f(\mathbf{x}_k) - f(\mathbf{x}_{k-1}) \leqslant \varepsilon|\beta_{\mathbf{x}k}|, \quad k \in [n]. \tag{A12}$$

Summing over k yields

$$f(\mathbf{x} + \varepsilon\boldsymbol{\beta}_{\mathbf{x}}) - f(\mathbf{x}) \leqslant \varepsilon \sum_{k=1}^{n} |\beta_{\mathbf{x}k}| = \varepsilon.$$

By (A11), the equality is attained, and hence each inequality in (A12) must be an equality. Take $i = \min\{k \in [n] : \beta_{\mathbf{x}k} \neq 0\}$. Then we have $\mathbf{x}_{i-1} = \mathbf{x}$ and $f(\mathbf{x}_i) - f(\mathbf{x}) = \varepsilon|\beta_{\mathbf{x}i}|$ with $\mathbf{x}_i - \mathbf{x} = \varepsilon\beta_{\mathbf{x}i}\mathbf{e}_i$. Since $\varepsilon > 0$ is arbitrary, we have there exists $i \in [n]$ such that

$$f(\mathbf{x} + \varepsilon\tilde{\mathbf{e}}_i) - f(\mathbf{x}) = \varepsilon, \quad \forall\, \varepsilon > 0,$$

where $\tilde{\mathbf{e}}_i \in \{\pm\mathbf{e}_i\}, i \in [n]$. By Lemma A2(*ii*), there also exists $\boldsymbol{\eta}_{\mathbf{x}}$ with $\|\boldsymbol{\eta}_{\mathbf{x}}\|_1 = 1$ such that $f(\mathbf{x}) - f(\mathbf{x} + \varepsilon\boldsymbol{\eta}_{\mathbf{x}}) = \varepsilon$, $\forall\, \varepsilon > 0$. Applying the same reasoning as above, we conclude that (A10) holds. This completes the proof. □

Proof of Proposition 1. Without loss of generality, assume that $c = 1$.

(a) It suffices to verify that f is 1-Lipschitz continuous and satisfies (A10). Denote by $I := \{i \in [n], \beta_i \neq 0\}$. For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{y}$,

$$\begin{aligned} f(\mathbf{x}) - f(\mathbf{y}) &= \boldsymbol{\beta}^\top(\mathbf{x} - \mathbf{y}) + g(\boldsymbol{\eta} \circ \mathbf{x}) - g(\boldsymbol{\eta} \circ \mathbf{y}) \\ &\leqslant \sum_{i \in I} |x_i - y_i| + \mathrm{Lip}(g) \sum_{i \in [n] \setminus I} |x_i - y_i| \leqslant \|\mathbf{x} - \mathbf{y}\|_1, \end{aligned}$$

where the first inequality follows from Hölder's inequality, and the second inequality follows from that $\mathrm{Lip}(g) \leqslant 1$. Thus f is 1-Lipschitz continuous. Next, for any $\mathbf{x} \in \mathbb{R}^n$ and $\varepsilon > 0$, choose $i \in I$ and define $\boldsymbol{v}_1 := \mathrm{sign}(\beta_i)\mathbf{e}_i$, where $\mathrm{sign}(\cdot)$ is the sign function. Then,

$$\begin{aligned} f(\mathbf{x} + \varepsilon\boldsymbol{v}_1) - f(\mathbf{x}) &= \boldsymbol{\beta}^\top(\mathbf{x} + \varepsilon\boldsymbol{v}_1) + g(\boldsymbol{\eta} \circ (\mathbf{x} + \varepsilon\boldsymbol{v}_1))) - \big(\boldsymbol{\beta}^\top\mathbf{x} + g(\boldsymbol{\eta} \circ \mathbf{x})\big) \\ &= \varepsilon\boldsymbol{\beta}^\top\boldsymbol{v}_1 + g(\boldsymbol{\eta} \circ \mathbf{x} + \varepsilon\boldsymbol{\eta} \circ \boldsymbol{v}_1) - g(\boldsymbol{\eta} \circ \mathbf{x}) = \varepsilon, \end{aligned}$$

where the last equality follows from $\boldsymbol{\beta}^\top\boldsymbol{v}_1 = \mathrm{sign}(\beta_i)\boldsymbol{\beta}^\top\mathbf{e}_i = |\beta_i| = 1$ and $\boldsymbol{\eta} \circ \boldsymbol{v}_1 = \mathrm{sign}(\beta_i)\boldsymbol{\eta} \circ \mathbf{e}_i = \mathbf{0}$ for the chosen $i \in I$, since $\boldsymbol{\beta} \circ \boldsymbol{\eta} = \mathbf{0}$. Similarly, for any $\mathbf{x} \in \mathbb{R}^n$ and $\varepsilon > 0$, by choosing $i \in I$ and $\boldsymbol{v}_2 := -\mathrm{sign}(\beta_i)\mathbf{e}_i$, one can verify that $f(\mathbf{x}) - f(\mathbf{x} + \varepsilon\boldsymbol{v}_2) = \varepsilon$. Hence, f satisfies (A10) with $\mathrm{Lip}(f) = 1$.

(b) It is straightforward to verify that f is 1-Lipschitz continuous. It remains to verify that f satisfies (A10). We only prove that for any $\mathbf{x} \in \mathbb{R}^n$, there exists $\tilde{\mathbf{e}}_i \in \{\pm\mathbf{e}_i\}$ for some $i \in [n]$ such that

$$f(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i)-f(\mathbf{x})=\varepsilon,\quad \forall\,\varepsilon>0 \tag{A13}$$

as the existence of $\tilde{\mathbf{e}}_j$ can be verified by applying the same argument to $-f$. Denote by $I_1=\{i\in[n],\beta_i\neq 0\}$. We show it by considering the following two cases.

(i) If $\boldsymbol{\beta}^\top\mathbf{x}\geqslant 0$, then by taking $\tilde{\mathbf{e}}_i=\mathrm{sign}(\beta_i)\mathbf{e}_i$ for some $i\in I_1$, we have

$$\begin{aligned}f(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i)-f(\mathbf{x}) &= |\boldsymbol{\beta}^\top(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i)|-|\boldsymbol{\nu}^\top(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i)|+g(\boldsymbol{\eta}\circ(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i))-f(\mathbf{x})\\ &\quad -|\boldsymbol{\beta}^\top\mathbf{x}|+|\boldsymbol{\nu}^\top\mathbf{x}|-g(\boldsymbol{\eta}\circ\mathbf{x})\\ &=\varepsilon-|\boldsymbol{\nu}^\top(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i)|+|\boldsymbol{\nu}^\top\mathbf{x}|+g(\boldsymbol{\eta}\circ(\mathbf{x}+\varepsilon\tilde{\mathbf{e}}_i))-g(\boldsymbol{\eta}\circ\mathbf{x})\\ &=\varepsilon,\end{aligned}$$

where the second equality follows from that $|\boldsymbol{\beta}^\top\mathbf{x}|=\boldsymbol{\beta}^\top\mathbf{x}$ and $\boldsymbol{\beta}^\top\tilde{\mathbf{e}}_i=\mathrm{sign}(\beta_i)\boldsymbol{\beta}^\top\mathbf{e}_i=|\beta_i|=1$, and the last equality follows from $\boldsymbol{\nu}^\top\tilde{\mathbf{e}}_i=0$ and $\boldsymbol{\eta}\circ\tilde{\mathbf{e}}_i=\mathbf{0}$ by $\boldsymbol{\beta}\circ\boldsymbol{\eta}=\boldsymbol{\nu}\circ\boldsymbol{\eta}=\mathbf{0}$. That is, (A13) holds.

(ii) If $\boldsymbol{\beta}^\top\mathbf{x}<0$, then by choosing $\tilde{\mathbf{e}}_i:=-\mathrm{sign}(\beta_i)\mathbf{e}_i$ for some $i\in I_1$, one can similarly prove (A13) holds.

Combining the above cases, we complete the proof. □

Proof of Proposition 2. Without loss of generality, assume that $\mathrm{Lip}(f)=1$. It suffices to show that (9) holds whenever f satisfies (8). We prove it by contradiction. Suppose there exist $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ such that $f(\mathbf{x})\leqslant f(\mathbf{y})$ and $\|\boldsymbol{\eta}_\mathbf{x}-\boldsymbol{\beta}_\mathbf{y}\|<\|\boldsymbol{\eta}_\mathbf{x}\|+\|\boldsymbol{\beta}_\mathbf{y}\|=2$. Define $\mathbf{z}=\mathbf{x}+c\boldsymbol{\eta}_\mathbf{x}$ and $\mathbf{z}_1=\mathbf{y}+c\boldsymbol{\beta}_\mathbf{y}$. There exists c large enough such that $\|\mathbf{x}-\mathbf{y}\|<c(2-\|\boldsymbol{\beta}_\mathbf{y}-\boldsymbol{\eta}_\mathbf{x}\|)$, and thus,

$$c_0:=\|\mathbf{z}_1-\mathbf{z}\|\leqslant\|\mathbf{x}-\mathbf{y}\|+c\|\boldsymbol{\beta}_\mathbf{y}-\boldsymbol{\eta}_\mathbf{x}\|<2c.$$

This implies that

$$\sup_{\|\mathbf{w}\|\leqslant c_0} f(\mathbf{z}+\mathbf{w})-f(\mathbf{z})\geqslant f(\mathbf{z}_1)-f(\mathbf{z})\geqslant f(\mathbf{z}_1)-f(\mathbf{y})+f(\mathbf{x})-f(\mathbf{z})=2c,$$

where the second inequality follows from $f(\mathbf{y})\geqslant f(\mathbf{x})$, and the equality follows from Lemma A2. This yields a contradiction to (8). Thus, we have for any $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$, if $f(\mathbf{x})\leqslant f(\mathbf{y})$ (hence including $\mathbf{x}=\mathbf{y}$), then $\|\boldsymbol{\eta}_\mathbf{x}-\boldsymbol{\beta}_\mathbf{y}\|=\|\boldsymbol{\eta}_\mathbf{x}\|+\|\boldsymbol{\beta}_\mathbf{y}\|$. Since $\|\boldsymbol{\eta}_\mathbf{x}\|=\|\boldsymbol{\beta}_\mathbf{y}\|=1$ and the norm is strictly convex, it follows that $\boldsymbol{\beta}_\mathbf{x}=-\boldsymbol{\eta}_\mathbf{x}=\boldsymbol{\beta}_\mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}\in\mathbb{R}^n$ with $f(\mathbf{x})\leqslant f(\mathbf{y})$. Otherwise we have

$\|\boldsymbol{\eta}_{\mathbf{x}}-\boldsymbol{\beta}_{\mathbf{y}}\|<2$, contradicting the requirement $\|\boldsymbol{\eta}_{\mathbf{x}}-\boldsymbol{\beta}_{\mathbf{y}}\|=\|\boldsymbol{\eta}_{\mathbf{x}}\|+\|\boldsymbol{\beta}_{\mathbf{y}}\|=2$. Therefore, we have $\boldsymbol{\beta}_{\mathbf{x}}$ can be chosen as the same $\boldsymbol{\beta}$ for all $\mathbf{x}\in\mathbb{R}^n$. It then follows that there exists $\boldsymbol{v}\in\mathbb{R}^n$ with $\|\boldsymbol{v}\|=1$ such that

$$f(\mathbf{x})-f(\mathbf{x}-\varepsilon\boldsymbol{v})=f(\mathbf{x}+\varepsilon\boldsymbol{v})-f(\mathbf{x})=\varepsilon,\quad \forall\,\mathbf{x}\in\mathbb{R}^n,\,\varepsilon>0.$$

Since the case $\varepsilon=0$ is trivially satisfied, it implies that (9) holds for any $\mathbf{x}\in\mathbb{R}^n$ and $t\in\mathbb{R}$.

For $\|\cdot\|=\|\cdot\|_a$, $a\in(1,\infty)$, it suffices to show for any $\mathbf{x}$, $\mathbf{y}\in\mathbb{R}^n$,

$$f(\mathbf{y})-f(\mathbf{x})=\boldsymbol{\beta}^\top(\mathbf{y}-\mathbf{x}), \tag{A14}$$

where $\boldsymbol{\beta}\in\mathbb{R}^n$ is the unique vector satisfying $\boldsymbol{\beta}^\top\boldsymbol{v}=\|\boldsymbol{\beta}\|_*\|\boldsymbol{v}\|=1$, that is, $\operatorname{sign}(\beta_i)=\operatorname{sign}(v_i)$ for $i\in[n]$ and $|\beta_i|^b=k|v_i|^a$ for some $k>0$ and b is the conjugate constant of a. By (9), we know (A14) holds for the case $\mathbf{y}-\mathbf{x}=t\boldsymbol{v}$, $t\in\mathbb{R}$. By $\boldsymbol{\beta}^\top\boldsymbol{v}=1$, we have any vector in $\mathbb{R}^n$ can be written as a linear combination of $\boldsymbol{v}$ and a vector $\boldsymbol{u}$ with $\boldsymbol{\beta}^\top\boldsymbol{u}=0$ and thus, it suffices to show (A14) for any $\mathbf{y}-\mathbf{x}=t\boldsymbol{u}$ with $\boldsymbol{\beta}^\top\boldsymbol{u}=0$. We show it by contradiction. Suppose that there exist $\mathbf{y}$ and $\mathbf{x}$ such that $\mathbf{y}-\mathbf{x}=\boldsymbol{u}$, $\boldsymbol{\beta}^\top\boldsymbol{u}=0$, and $f(\mathbf{x})<f(\mathbf{y})$. Denote by $\varepsilon=f(\mathbf{y})-f(\mathbf{x})>0$. Then we have

$$f(\mathbf{y}+t\boldsymbol{v})-f(\mathbf{x})=f(\mathbf{y}+t\boldsymbol{v})-f(\mathbf{y})+f(\mathbf{y})-f(\mathbf{x})=t+\varepsilon,$$

and

$$\|\mathbf{y}+t\boldsymbol{v}-\mathbf{x}\|^a=\|\boldsymbol{u}+t\boldsymbol{v}\|^a=\sum_{i=1}^n|u_i+tv_i|^a.$$

It follows that

$$\frac{f(\mathbf{y}+t\boldsymbol{v})-f(\mathbf{x})}{\|\mathbf{y}+t\boldsymbol{v}-\mathbf{x}\|}=\frac{t+\varepsilon}{\left(\sum_{i=1}^n|u_i+tv_i|^a\right)^{1/a}}=\frac{1+s\varepsilon}{\left(\sum_{i=1}^n|su_i+v_i|^a\right)^{1/a}}, \tag{A15}$$

where $s=1/t$. Denote by $I_1=\{i:v_i>0\}$, $I_2=\{i:v_i<0\}$, $I_3=\{i:v_i=0,u_i\neq0\}$, and $g(s)=(1+s\varepsilon)^a-\sum_{i=1}^n|su_i+v_i|^a$. We have as $s\downarrow0$,

$$\begin{aligned}
g'(s)&=a(1+s\varepsilon)^{a-1}\varepsilon-\sum_{i\in I_1}a(su_i+v_i)^{a-1}u_i+\sum_{i\in I_2}a(-su_i-v_i)^{a-1}u_i-\sum_{i\in I_3}as^{a-1}|u_i|^a\\
&\overset{\text{sign}}{=}(1+s\varepsilon)^{a-1}\varepsilon-\sum_{i\in I_1}(su_i+v_i)^{a-1}u_i+\sum_{i\in I_2}(-su_i-v_i)^{a-1}u_i-\sum_{i\in I_3}s^{a-1}|u_i|^a\\
&\to\varepsilon-\sum_{i\in I_1}v_i^{a-1}u_i+\sum_{i\in I_2}(-v_i)^{a-1}u_i=\varepsilon-\sum_{i=1}^n\operatorname{sign}(v_i)|v_i|^{a-1}u_i=\varepsilon>0,
\end{aligned}$$

where the last equality follows from that $\boldsymbol{\beta}^\top \boldsymbol{u} = 0$, $|\beta_i| = k_1 |v_i|^{a-1}$ and $\mathrm{sign}(v_i) = \mathrm{sign}(\beta_i)$, $i \in [n]$. This, together with $g(0) = 0$, implies $g(s) > 0$ for some $s > 0$. Substituting this into (A15) yields that $\frac{f(\mathbf{y}+t\boldsymbol{v})-f(\mathbf{x})}{\|\mathbf{y}+t\boldsymbol{v}-\mathbf{x}\|} > 1$ for some $t > 0$. This yields a contradiction to (9). Therefore, we have (A14) holds and thus we complete the proof. □

Proof of Theorem 2. For $(i) \Rightarrow (ii)$, we can employ the exact arguments to prove (A3) in the proof for the direction $(i) \Rightarrow (iii)$ of Theorem 1 where we choose $\rho = \mathrm{VaR}_\alpha$ for some $\alpha \in [0, 1)$ which is a monotone risk measure to prove (A3) which is exactly the statement in (ii). It remains to show $(ii) \Rightarrow (i)$. Note that if $\mathrm{Lip}(f) = 0$, then f is a constant function. Taking $c_f = 0$, (10) holds trivially. We next consider $\mathrm{Lip}(f) > 0$. Assume without loss of generality that $\mathrm{Lip}(f) = 1$. We aim to show that (10) holds for $c_f := 1$. First note that for any random vector $\boldsymbol{\xi}$ with $\mathbb{E}[\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p] \leqslant \varepsilon^p$, by 1-Lipschitz continuity of f, we have

$$\mathbb{E}\left[|f(\boldsymbol{\xi}) - f(\boldsymbol{\zeta})|^p\right] \leqslant \mathbb{E}[\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p] \leqslant \varepsilon^p,$$

and thus, $\{F_{f(\boldsymbol{\xi})} : F_{\boldsymbol{\xi}} \in \mathbb{B}_p(F_0, \varepsilon)\} \subseteq \mathcal{C}_p(f|F_0, \varepsilon)$. It follows that

$$\sup_{F \in \mathbb{B}_p(F_0,\varepsilon)} \rho^F\left(f(\boldsymbol{\xi})\right) \leqslant \sup_{G \in \mathcal{C}_p(f|F_0,\varepsilon)} \rho^G(X). \tag{A16}$$

We next show the reverse direction. It suffices to demonstrate that for any $G \in \mathcal{C}_p(f|F_0, \varepsilon)$ and $Z \sim G$ with $\mathbb{E}[|Z - f(\boldsymbol{\zeta})|^p] \leqslant \varepsilon^p$, there exists $F_{\boldsymbol{\xi}} \in \mathbb{B}_p(F_0, \varepsilon)$ and $\boldsymbol{\xi} \sim F_{\boldsymbol{\xi}}$ such that

$$\rho^{F_{\boldsymbol{\xi}}}(f(\boldsymbol{\xi})) \geqslant \rho^G(Z).$$

To this end, take $G \in \mathcal{C}_p(f|F_0, \varepsilon)$ and $Z \sim G$. Define $Z^* = \max\{f(\boldsymbol{\zeta}), Z\}$ and denote by G^* the distribution of Z^*. We have $\rho^G(Z) \leqslant \rho^{G^*}(Z^*)$ and $\mathbb{E}[|Z^* - f(\boldsymbol{\zeta})|^p] \leqslant \mathbb{E}[|Z - f(\boldsymbol{\zeta})|^p] \leqslant \varepsilon^p$, and thus, $G^* \in \mathcal{C}_p(f|F_0, \varepsilon)$. So, without loss of generality, assume that $Z \geqslant f(\boldsymbol{\zeta})$ almost surely. Denote by $T := Z - f(\boldsymbol{\zeta})$. We have $T \geqslant 0$ almost surely and $\mathbb{E}[T^p] \leqslant \varepsilon^p$. With similar arguments as in the proof of Theorem 1 $(iii) \Rightarrow (ii)$, there exists a measurable mapping $\mathbf{V}$ such that

$$\mathbf{V}(\omega) \in \operatorname*{arg\,max}_{\|\mathbf{y}\| \leqslant T(\omega)} f\left(\boldsymbol{\zeta}(\omega) + \mathbf{y}\right), \quad \omega \in \Omega.$$

Define $\boldsymbol{\xi}(\omega) = \boldsymbol{\zeta}(\omega) + \mathbf{V}(\omega)$, $\omega \in \Omega$. It follows that for any $\omega \in \Omega$,

$$f(\boldsymbol{\xi}(\omega)) = f(\boldsymbol{\zeta}(\omega) + \mathbf{V}(\omega)) = \max_{\|\mathbf{y}\| \leqslant T(\omega)} f(\boldsymbol{\zeta}(\omega) + \mathbf{y}) = f(\boldsymbol{\zeta}(\omega)) + T(\omega) = Z(\omega).$$

where the second equality follows from the definition of $\mathbf{V}$, the third equality follows from (11) and the last one follows from the deifnition of T. Moreover, noting that $\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|^p] = \mathbb{E}[\|\mathbf{V}\|^p] \leqslant \mathbb{E}[T^p] \leqslant \varepsilon^p$, we have $F_{\boldsymbol{\xi}} \in \mathbb{B}_p(F_0,\varepsilon)$. Hence

$$\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \rho^F\left(f(\boldsymbol{\xi})\right) \geqslant \sup_{G\in\mathcal{C}_p(f|F_0,\varepsilon)} \rho^G(X). \tag{A17}$$

Combining (A17) and (A16), we have (10) holds, completing the proof. □

Proof of Proposition 3. Note that the case $\mathrm{Lip}(f) = 0$ is trivial. It suffices to consider the case $\mathrm{Lip}(f) > 0$. Without loss of generality, let f satisfy (11) with $\mathrm{Lip}(f) = 1$. We first show that for each $\mathbf{x} \in \mathbb{R}^n$, there exists $\boldsymbol{\eta}_{\mathbf{x}} \in \partial f(\mathbf{x})$ such that

$$\|\boldsymbol{\eta}_{\mathbf{x}}\|_* = 1, \tag{A18}$$

where $\partial f(\mathbf{x})$ denotes the subdifferential of f at $\mathbf{x}$. Note that by Lemma A2, for every $\mathbf{x} \in \mathbb{R}^n$, there exists $\boldsymbol{\beta}_{\mathbf{x}} \in \mathbb{R}^n$ with $\|\boldsymbol{\beta}_{\mathbf{x}}\| = 1$ such that

$$f(\mathbf{x}+\varepsilon\boldsymbol{\beta}_{\mathbf{x}}) - f(\mathbf{x}) = \varepsilon, \quad \forall\, \varepsilon > 0.$$

This implies that the directional derivative of f at $\mathbf{x}$ at direction $\boldsymbol{\beta}_{\mathbf{x}}$ equals to 1. Therefore, by Theorem 23.4 of Rockafellar (1970), there exists a subgradient $\boldsymbol{\eta}_{\mathbf{x}} \in \partial f(\mathbf{x})$ such that $\boldsymbol{\eta}_{\mathbf{x}}^\top \boldsymbol{\beta}_{\mathbf{x}} = 1$, and hence,

$$\boldsymbol{\eta}_{\mathbf{x}}^\top \boldsymbol{\beta}_{\mathbf{x}} = 1 \leqslant \|\boldsymbol{\eta}_{\mathbf{x}}\|_* \|\boldsymbol{\beta}_{\mathbf{x}}\| = \|\boldsymbol{\eta}_{\mathbf{x}}\|_*.$$

It then follows from $\|\boldsymbol{\eta}_{\mathbf{x}}\|_* \leqslant 1$ by the 1-Lipschitz continuity of f that (A18) holds. Note that a convex function f can be written as

$$f(\mathbf{x}) = \max_{\mathbf{z}\in\mathbb{R}^n} \{f(\mathbf{z}) + \boldsymbol{\eta}_{\mathbf{z}}^\top(\mathbf{x}-\mathbf{z})\}.$$

We have (12) holds and thus, we complete the proof. □

Proof of Proposition 4. The "if" part is a direct consequence of $(iii) \Rightarrow (i)$ in Theorem 2 and the "only if" part has been already proved in the argument for $(i) \Rightarrow (iii)$ in Theorem 2. □

Proof of Lemma 1. Note that (14) holds, that is,

$$\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \mathrm{VaR}_\alpha^F\left(\max_{i\in I} \frac{\boldsymbol{\beta}_i^\top \boldsymbol{\xi}}{\|\boldsymbol{\beta}_i\|_*}\right) = \sup_{G\in\mathbb{B}_p(G_0,\varepsilon)} \mathrm{VaR}_\alpha^G(X) =: \overline{\mathrm{VaR}}_\alpha(G_0),$$

where $G_0 \in \mathcal{M}(\mathbb{R})$ is the distribution of $\max_{i\in I} \boldsymbol{\beta}_i^\top \boldsymbol{\zeta}/\|\boldsymbol{\beta}_i\|_*$ and $\boldsymbol{\zeta} \sim F_0$. Hence, it suffices to show $\overline{\mathrm{VaR}}_\alpha(G_0)$ is the unique solution of x to (15). By Proposition 4 of Liu et al. (2022), we have that $\overline{\mathrm{VaR}}_\alpha(G_0)$ equals to the unique solution to

$$\int_\alpha^1 (x - \mathrm{VaR}_u^{G_0}(X))_+^p \mathrm{d}u = \varepsilon^p. \tag{A19}$$

Note that $\mathrm{VaR}_u(h(X)) = h(\mathrm{VaR}_u(X))$ for $h(t) = -(x-t)_+^p$ which is an increasing and continuous function in $t \in \mathbb{R}$. It follows that (A19) is equivalent to

$$\int_\alpha^1 -\mathrm{VaR}_u^{G_0}\left(-(x-X)_+^p\right) \mathrm{d}u = \varepsilon^p,$$

or equivalently,

$$\mathrm{CVaR}_\alpha^{G_0}\left(-(x-X)_+^p\right) = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u^{G_0}\left(-(x-X)_+^p\right) \mathrm{d}u = -\frac{\varepsilon^p}{1-\alpha},$$

Therefore, $\overline{\mathrm{VaR}}_\alpha(G_0)$ is the unique solution of x to (15), which completes the proof. □

To prove Theorem 3, we need the following lemma. We call a function f continuous, if it is continuous with respect to the distance induced by the norm $\|\cdot\|$.

Lemma A4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. If there exist $\mathbf{x}_0 \in \mathbb{R}^n$ and $\varepsilon_0 > 0$ such that*

$$k := \sup_{\|\mathbf{y}\|\leqslant\varepsilon_0} \frac{f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0)}{\varepsilon_0} < 1, \tag{A20}$$

then for any $k_1 \in (k, 1)$, there exist $\mathbf{x}_1 \in \mathbb{R}^n$ and $\varepsilon_1 > 0$ such that

$$f(\mathbf{x}_1 + \mathbf{y}) - f(\mathbf{x}_1) \leqslant k_1\|\mathbf{y}\|, \quad \forall\, \mathbf{y} \text{ with } \|\mathbf{y}\| \leqslant \varepsilon_1.$$

Proof. For $k_1 \in (k, 1)$, define $g(\mathbf{y}) = f(\mathbf{x}_0 + \mathbf{y}) - k_1\|\mathbf{y}\|$, $\mathbf{y} \in \mathbb{R}^n$, which continuous as f is continuous. Noting that $\{\mathbf{y} : \|\mathbf{y}\| \leqslant \varepsilon_0\}$ is compact, there exists $\mathbf{y}^*$ with $\|\mathbf{y}^*\| \leqslant \varepsilon_0$ such that $g(\mathbf{y}^*) = \max_{\|\mathbf{y}\|\leqslant\varepsilon_0} g(\mathbf{y})$. Moreover, note that for any $\mathbf{y}$ with $\|\mathbf{y}\| = \varepsilon_0$, we have

$$g(\mathbf{y}) = f(\mathbf{x}_0 + \mathbf{y}) - k_1\varepsilon_0 \leqslant f(\mathbf{x}_0) + k\varepsilon_0 - k_1\varepsilon_0 < f(\mathbf{x}_0) = g(\mathbf{0}) \leqslant g(\mathbf{y}^*),$$

where the first inequality follows from the definition of k and the strict one from $k < k_1$. Therefore, we have $\|\mathbf{y}^*\| < \varepsilon_0$. Define

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{y}^* \quad \text{and} \quad \varepsilon_1 = \varepsilon_0 - \|\mathbf{y}^*\|.$$

For any $\mathbf{y}$ with $\|\mathbf{y}\| \leqslant \varepsilon_1$, we have $\|\mathbf{y}+\mathbf{y}^*\| \leqslant \|\mathbf{y}\|+\|\mathbf{y}^*\| \leqslant \varepsilon_1+\|\mathbf{y}^*\| = \varepsilon_0$. Then, by the definition of $\mathbf{y}^*$, we have

$$f(\mathbf{x}_0+\mathbf{y}^*+\mathbf{y}) - k_1\|\mathbf{y}^*+\mathbf{y}\| = g(\mathbf{y}+\mathbf{y}^*) \leqslant g(\mathbf{y}^*) = f(\mathbf{x}_0+\mathbf{y}^*) - k_1\|\mathbf{y}^*\|,$$

and hence

$$f(\mathbf{x}_1+\mathbf{y}) - f(\mathbf{x}_1) \leqslant k_1(\|\mathbf{y}^*+\mathbf{y}\| - \|\mathbf{y}^*\|) \leqslant k_1\|\mathbf{y}\|, \quad \forall\, \|\mathbf{y}\| \leqslant \varepsilon_1.$$

This completes the proof. □

Proof of Theorem 3. The "if" part is a direct consequence of $(i) \Rightarrow (ii)$ in Theorem 2. For the "only if" part, suppose that there exists $c_f \geqslant 0$ such that (19) holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$. If $c_f = 0$, then (19) reduces to

$$\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \rho^F(f(\boldsymbol{\xi})) = \sup_{G\in\mathcal{C}_p(f|F_0,0)} \rho^G(X) = \rho^{F_0}(f(\boldsymbol{\zeta})). \tag{A21}$$

We show that f is a constant function by contradiction. Suppose there exist $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$ such that $f(\mathbf{x}) < f(\mathbf{y})$. Take $F_0 = \delta_{\mathbf{x}}$ and $\varepsilon = \|\mathbf{x}-\mathbf{y}\|$. Then we have $\delta_{\mathbf{y}} \in \mathbb{B}_p(F_0,\varepsilon)$ and thus,

$$\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \rho^F(f(\boldsymbol{\xi})) \geqslant \rho^{\delta_{\mathbf{y}}}(f(\boldsymbol{\xi})) = f(\mathbf{y}) > f(\mathbf{x}) = \rho^{F_0}(f(\boldsymbol{\zeta})),$$

which contradicts (A21). Therefore, f is a constant function and thus satisfies (11). Now consider the case $c_f > 0$. By the positive homogeneity of ρ, we assume without loss of generality that $c_f = 1$. For $\rho = \mathbb{E}$, equation (19) then becomes

$$\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p} \mathbb{E}[f(\boldsymbol{\xi}+\boldsymbol{\zeta})] = \sup_{\mathbb{E}|X|^p\leqslant\varepsilon^p} \mathbb{E}[X+f(\boldsymbol{\zeta})] = \mathbb{E}^{F_0}[f(\boldsymbol{\zeta})] + \varepsilon, \tag{A22}$$

where $\boldsymbol{\zeta} \sim F_0$, the second equality follows from Theorem 3 in Wu et al. (2022). We next show that (A22) implies

$$\sup_{\|\mathbf{y}\|\leqslant\varepsilon} f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) = \varepsilon, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\, \varepsilon > 0. \tag{A23}$$

To see it, first note that for any $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$, by setting $\boldsymbol{\zeta} \sim F_0 := \delta_{\mathbf{x}}$ and $\varepsilon := \|\mathbf{y}-\mathbf{x}\|$, we have (A22) implies

$$f(\mathbf{y}) \leqslant \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p} \mathbb{E}[f(\boldsymbol{\xi}+\mathbf{x})] = f(\mathbf{x}) + \varepsilon = f(\mathbf{x}) + \|\mathbf{y}-\mathbf{x}\|.$$

It follows that f is Lipschitz continuous with $\mathrm{Lip}(f) \leqslant 1$. Thus, we have

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) \leqslant \varepsilon, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\, \varepsilon > 0.$$

We next show that f satisfies (A23) by contradiction. Suppose, for contradiction, that there exist $\mathbf{x}_0 \in \mathbb{R}^n$ and $\varepsilon_0 > 0$ such that

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} f(\mathbf{x}_0+\mathbf{y}) - f(\mathbf{x}_0) < \varepsilon_0. \tag{A24}$$

We next consider two cases: $p = \infty$ and $p \in (1, \infty)$. When $p = \infty$, we have

$$\begin{aligned} \sup_{\text{ess-sup}(\|\boldsymbol{\xi}\|) \leqslant \varepsilon_0} \mathbb{E}[f(\mathbf{x_0}+\boldsymbol{\xi})] &\leqslant \mathbb{E}\left[\sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} f(\mathbf{x}_0+\mathbf{y})\right] \\ &= \sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} f(\mathbf{x}_0+\mathbf{y}) < f(\mathbf{x}_0) + \varepsilon_0, \end{aligned}$$

where the first inequality follows from the monotonicity of $\mathbb{E}$ and $f(\mathbf{x}_0+\boldsymbol{\xi}) \leqslant \sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} f(\mathbf{x}_0+\mathbf{y})$ almost surely for any $\boldsymbol{\xi}$ with $\text{ess-sup}(\|\boldsymbol{\xi}\|) \leqslant \varepsilon_0$. This contradicts (A22). Now suppose $p \in (1, \infty)$. Define

$$k := \sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} \left\{ \frac{f(\mathbf{x}_0+\mathbf{y}) - f(\mathbf{x}_0)}{\varepsilon_0} \right\}.$$

By (A24), we have $k < 1$. By Lemma A4, there exist $k_1 \in (k, 1)$, $\mathbf{x}_1 \in \mathbb{R}^n$ and $\varepsilon_1 > 0$ such that

$$f(\mathbf{x}_1+\mathbf{y}) - f(\mathbf{x}_1) \leqslant k_1 \|\mathbf{y}\|, \quad \forall\, \|\mathbf{y}\| \leqslant \varepsilon_1.$$

For sufficiently small $\varepsilon < \varepsilon_1$,

$$
\begin{aligned}
&\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \mathbb{E}[f(\mathbf{x}_1 + \boldsymbol{\xi})] \\
&= \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \mathbb{E}\left[f(\mathbf{x}_1 + \boldsymbol{\xi}) \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant \varepsilon_1\}} + f(\mathbf{x}_1 + \boldsymbol{\xi}) \mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_1\}}\right] \\
&\leqslant \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \mathbb{E}\left[(f(\mathbf{x}_1) + k_1\|\boldsymbol{\xi}\|) \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant \varepsilon_1\}} + (f(\mathbf{x}_1) + \|\boldsymbol{\xi}\|) \mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_1\}}\right] \\
&= f(\mathbf{x}_1) + \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \mathbb{E}\left[k_1\|\boldsymbol{\xi}\| + (1-k_1)\|\boldsymbol{\xi}\| \mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_1\}}\right] \\
&\leqslant f(\mathbf{x}_1) + \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} k_1 \mathbb{E}[\|\boldsymbol{\xi}\|] + \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} (1-k_1)\mathbb{E}[\|\boldsymbol{\xi}\| \mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_1\}}] \\
&\leqslant f(\mathbf{x}_1) + k_1\varepsilon + \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} (1-k_1)\mathbb{E}[\|\boldsymbol{\xi}\| \mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_1\}}] \\
&\leqslant f(\mathbf{x}_1) + k_1\varepsilon + \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} (1-k_1)(\mathbb{E}[\|\boldsymbol{\xi}\|^p])^{1/p} (\mathbb{E}\mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_1\}}])^{1/q} \\
&\leqslant f(\mathbf{x}_1) + k_1\varepsilon + (1-k_1)\varepsilon \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} (\mathbb{E}\mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_1\}}])^{1/q} \\
&\leqslant f(\mathbf{x}_1) + k_1\varepsilon + (1-k_1)\varepsilon \left(\frac{\varepsilon}{\varepsilon_1}\right)^{p/q} < f(\mathbf{x}_1) + \varepsilon,
\end{aligned} \tag{A25}
$$

where the first inequality follows from that f is Lipschitz continuous with $\mathrm{Lip}(f) \leqslant 1$, the third inequality follows from $\mathbb{E}[\|\boldsymbol{\xi}\|] \leqslant (\mathbb{E}[\|\boldsymbol{\xi}\|^p])^{1/p}$, the fourth inequality follows from Hölder's inequality and $k_1 < 1$, the sixth inequality follows from Markov's inequality, and the strict inequality holds because $\varepsilon < \varepsilon_1$ and $p/q = p - 1 > 0$. Then, (A25) leads to a contradiction to (A22), thereby establishing (A23). This completes the proof. □

Proof of Proposition 5. The proof follows directly from that of Theorem 3. □

To show Theorem 4, we need the following lemma.

Lemma A5. *For $\alpha \in [0, 1)$ and $\varepsilon > 0$, if f is a function satisfying the condition of Theorem 4(i), then for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$*

$$
\begin{aligned}
\sup_{F \in \mathbb{B}_1(F_0, \varepsilon)} \mathrm{CVaR}_\alpha^F(f(\boldsymbol{\xi})) &= \sup_{G \in \mathcal{C}_1(f|F_0, c_f\varepsilon)} \mathrm{CVaR}_\alpha^G(X) \\
&= \mathrm{CVaR}_\alpha^{F_0}(f(\boldsymbol{\zeta})) + \frac{\varepsilon c_f}{1-\alpha},
\end{aligned} \tag{A26}
$$

where $\boldsymbol{\zeta} \sim F_0$ and $c_f = \mathrm{Lip}(f)$.

Proof. Assume without loss generality that $\mathrm{Lip}(f) = 1$. Note that the second equality in (A26) has been given in Proposition 2 in Pflug et al. (2012) and one can easily check that $f_\# \mathbb{B}_1(F_0, \varepsilon) \subseteq$

$\mathcal{C}_1(f|F_0, \varepsilon)$ when f is 1-Lipschitz continuous. It suffices to show

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|]\leqslant\varepsilon} \mathrm{CVaR}_\alpha\left(f\left(\boldsymbol{\xi}\right)\right) \geqslant \mathrm{CVaR}_\alpha^{F_0}(f(\boldsymbol{\zeta})) + \frac{\varepsilon}{1-\alpha}. \tag{A27}$$

Denote by $s_k = \limsup_{m\to\infty}(f(\mathbf{x}_0 + m\boldsymbol{v}_k) - f(\mathbf{x}_0))/m$. By (20), we have $\lim_{k\to\infty} s_k = 1$, and thus, there exists a sequence $\{k_j\}_{j\in\mathbb{N}}$ such that $s_{k_j} > 1 - 1/2j$, $i \in \mathbb{N}$. By the definition of s_{k_j}, there exists a sequence $\{m_j\}_{j\in\mathbb{N}}$ such that $m_j \to \infty$ as $j \to \infty$ and

$$\frac{f(\mathbf{x}_0 + m_j\boldsymbol{v}_{k_j}) - f(\mathbf{x}_0)}{m_j} \geqslant 1 - \frac{1}{j}, \quad \forall j \in \mathbb{N}.$$

It then follows that for any $\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned} f(\mathbf{x} + m_j\boldsymbol{v}_{k_j}) - f(\mathbf{x}) &\geqslant f(\mathbf{x}_0 + m_j\boldsymbol{v}_{k_j}) - f(\mathbf{x}_0) - 2\|\mathbf{x} - \mathbf{x}_0\| \\ &\geqslant \tilde{m}_j - 2\|\mathbf{x} - \mathbf{x}_0\|, \end{aligned} \tag{A28}$$

where $\tilde{m}_j := m_j(1 - 1/j)$ and the first inequality uses the fact that f is 1-Lipschitz continuous. Denote by U a uniform random variable on $[0,1]$ such that U and $f(\boldsymbol{\zeta})$ are comonotonic.[4] Define

$$\boldsymbol{\xi}_j = m_j\boldsymbol{v}_{k_j}\mathbb{1}_{A_j} + \boldsymbol{\zeta}, \quad j \in \mathbb{N}, \tag{A29}$$

where $A_j := \{1 - \varepsilon/m_j < U \leqslant 1\}$. One can verify that $\mathbb{E}\left[\|\boldsymbol{\xi}_j - \boldsymbol{\zeta}\|\right] = \varepsilon$. Then, we have

$$\begin{aligned} \mathrm{CVaR}_\alpha(f(\boldsymbol{\xi}_j)) - \mathrm{CVaR}_\alpha(f(\boldsymbol{\zeta})) &\geqslant \mathrm{CVaR}_\alpha\left(f(\boldsymbol{\xi}_j) - f(\boldsymbol{\zeta})\right) \\ &= \mathrm{CVaR}_\alpha\left((f(\boldsymbol{\zeta} + m_j\boldsymbol{v}_{k_j}) - f(\boldsymbol{\zeta}))\mathbb{1}_{A_j}\right) \\ &\geqslant \mathrm{CVaR}_\alpha\left((\tilde{m}_j - 2\|\boldsymbol{\zeta} - \mathbf{x}_0\|)\mathbb{1}_{A_j}\right) \\ &\geqslant \mathrm{CVaR}_\alpha\left(\tilde{m}_j\mathbb{1}_{A_j}\right) - \mathrm{CVaR}_\alpha\left(2\|\boldsymbol{\zeta} - \mathbf{x}_0\|\mathbb{1}_{A_j}\right) \\ &\geqslant \tilde{m}_j\,\mathrm{CVaR}_\alpha\left(\mathbb{1}_{A_j}\right) - \frac{2}{1-\alpha}\mathbb{E}\left[\|\boldsymbol{\zeta} - \mathbf{x}_0\|\mathbb{1}_{A_j}\right] \\ &\to \frac{\varepsilon}{1-\alpha} \quad \text{as } j \to \infty, \end{aligned}$$

where the first and the third inequalities follow from the subadditivity of CVaR_α, the second inequality follows from (A28), the last inequality follows from $\mathrm{CVaR}_\alpha(X) \leqslant \mathbb{E}[X]/(1-\alpha)$ for any nonnegative random variable X, and the last limit follows from $\mathrm{CVaR}_\alpha(\mathbb{1}_{A_j}) = \varepsilon/(m_j(1-\alpha))$ for

[4] Two random variables X, Y are called comonotonic if there exist a random variable Z and two nondecreasing functions f, g such that $X = f(Z)$ and $Y = g(Z)$ almost surely.

$m_j > \varepsilon/(1-\alpha)$ and $\lim_{j\to\infty} \mathbb{E}[\|\boldsymbol{\zeta} - \mathbf{x}_0\| \mathbb{1}_{A_j}] = 0$ by the dominated convergence theorem. Thus, we have

$$\sup_{\mathbb{E}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\leqslant\varepsilon} \mathrm{CVaR}_\alpha\left(f\left(\boldsymbol{\xi}\right)\right) \geqslant \liminf_{j\to\infty} \mathrm{CVaR}_\alpha(f(\boldsymbol{\xi}_j)) \geqslant \mathrm{CVaR}_\alpha(f(\boldsymbol{\zeta})) + \frac{\varepsilon}{1-\alpha}.$$

That is, (A27) holds, which completes the proof. □

Proof of Theorem 4. (*i*) Without loss generality assume $\mathrm{Lip}(f) = 1$. Note that

$$\begin{aligned}
\sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \rho^F\left(f\left(\boldsymbol{\xi}\right)\right) &= \sup_{\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|]\leqslant\varepsilon} \sup_{\mu\in\mathscr{M}_\rho} \int_0^1 \mathrm{CVaR}_\alpha(f\left(\boldsymbol{\xi}\right))\mathrm{d}\mu(\alpha) \\
&= \sup_{\mu\in\mathscr{M}_\rho} \sup_{\mathbb{E}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\leqslant\varepsilon} \int_0^1 \mathrm{CVaR}_\alpha(f\left(\boldsymbol{\xi}\right))\mathrm{d}\mu(\alpha) \\
&\leqslant \sup_{\mu\in\mathscr{M}_\rho} \int_0^1 \sup_{\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|]\leqslant\varepsilon} \mathrm{CVaR}_\alpha(f\left(\boldsymbol{\xi}\right))\mathrm{d}\mu(\alpha) \\
&= \sup_{\mu\in\mathscr{M}_\rho} \int_0^1 \mathrm{CVaR}_\alpha^{F_0}(f(\boldsymbol{\zeta})) + \frac{\varepsilon}{1-\alpha}\mathrm{d}\mu(\alpha),
\end{aligned} \tag{A30}$$

where the third equality follows from Lemma A5. We next show that the inequality in (A30) is an equality. Define $\boldsymbol{\xi}_i$ by (A29) as in the proof of Lemma A5 which satisfies $\mathbb{E}[\|\boldsymbol{\xi}_j - \boldsymbol{\zeta}\|] \leqslant \varepsilon$ and

$$\liminf_{j\to\infty} \mathrm{CVaR}_\alpha(f(\boldsymbol{\xi}_j)) \geqslant \mathrm{CVaR}_\alpha(f(\boldsymbol{\zeta})) + \varepsilon/(1-\alpha).$$

Therefore,

$$\begin{aligned}
\sup_{\mu\in\mathscr{M}_\rho} \sup_{\mathbb{E}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|\leqslant\varepsilon} \int_0^1 \mathrm{CVaR}_\alpha(f\left(\boldsymbol{\xi}\right))\mathrm{d}\mu(\alpha) &\geqslant \sup_{\mu\in\mathscr{M}_\rho} \sup_{j\in\mathbb{N}} \int_0^1 \mathrm{CVaR}_\alpha(f\left(\boldsymbol{\xi}_j\right))\mathrm{d}\mu(\alpha) \\
&\geqslant \sup_{\mu\in\mathscr{M}_\rho} \int_0^1 \liminf_{j\to\infty} \mathrm{CVaR}_\alpha(f\left(\boldsymbol{\xi}_j\right))\mathrm{d}\mu(\alpha) \\
&\geqslant \sup_{\mu\in\mathscr{M}_\rho} \int_0^1 \mathrm{CVaR}_\alpha^{F_0}(f(\boldsymbol{\zeta})) + \frac{\varepsilon}{1-\alpha}\mathrm{d}\mu(\alpha),
\end{aligned}$$

where the second inequality follows from Fatou's Lemma. This means that the inequality of (A30) is actually an equality. That is,

$$\sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \rho^F\left(f\left(\boldsymbol{\xi}\right)\right) = \sup_{\mu\in\mathscr{M}_\rho} \int_0^1 \mathrm{CVaR}_\alpha^{F_0}(f(\boldsymbol{\zeta})) + \frac{\varepsilon}{1-\alpha}\mathrm{d}\mu(\alpha). \tag{A31}$$

Similarly, one can prove that

$$\begin{aligned}\sup_{G\in\mathcal{C}_1(f|F_0,\varepsilon)}\rho^G(X) &= \sup_{\mathbb{E}\|X-f(\boldsymbol{\zeta})\|\leqslant\varepsilon}\sup_{\mu\in\mathscr{M}_\rho}\int_0^1 \mathrm{CVaR}_\alpha(X)\mathrm{d}\mu(\alpha)\\ &= \sup_{\mu\in\mathscr{M}_\rho}\int_0^1 \mathrm{CVaR}_\alpha^{F_0}(f(\boldsymbol{\zeta}))+\frac{\varepsilon}{1-\alpha}\mathrm{d}\mu(\alpha).\end{aligned}\tag{A32}$$

Combining (A31) with (A32) yields (21), which completes the proof of (i).

(ii) We first show that f is Lipschitz continuous by contradiction. Suppose that f is not Lipschitz continuous. Then for any $n\in\mathbb{N}$, there exist $\mathbf{x}_n$, $\mathbf{y}_n$ such that $|f(\mathbf{x}_n)-f(\mathbf{y}_n)|>n\|\mathbf{x}_n-\mathbf{y}_n\|$. Without loss of generality assume $f(\mathbf{x}_n)-f(\mathbf{y}_n)>n\|\mathbf{x}_n-\mathbf{y}_n\|$, $n\in\mathbb{N}$. Take $F_n=\delta_{\mathbf{y}_n}$ and $\varepsilon_n=\|\mathbf{x}_n-\mathbf{y}_n\|$. It follows that

$$\sup_{F\in\mathbb{B}_1(F_n,\varepsilon_n)}\rho^F(f(\boldsymbol{\xi}))=\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\mathbf{y}_n\|]\leqslant\varepsilon_n}\rho(f(\boldsymbol{\xi}))\geqslant f(\mathbf{x}_n)>f(\mathbf{y}_n)+n\varepsilon_n.\tag{A33}$$

Note that

$$\sup_{G\in\mathcal{C}_1(f|F_n,c_f\varepsilon_n)}\rho^G(X)=\sup_{\mathbb{E}[|X-f(\mathbf{y}_n)|]\leqslant c_f\varepsilon_n}\rho(X)\leqslant f(\mathbf{y}_n)+C_\rho c_f\varepsilon_n,$$

where the first inequality follows from the Kuosuoka representation (5) and $C_\rho=\sup_{\mu\in\mathscr{M}_\rho}\int_0^1\frac{1}{1-\alpha}\mathrm{d}\mu(\alpha)$. This, together with (A33), yields a contradiction to (21) by noting n can be arbitrarily large. Therefore, f is a Lipschitz continuous function.

We next show that (20) holds. Without loss of generality, assume $\mathrm{Lip}(f)=1$. For $\mathbf{x}_0\in\mathbb{R}^n$ and any $\boldsymbol{v}\in\mathbb{R}^n$ with $\|\boldsymbol{v}\|=1$, define

$$\phi(\boldsymbol{v}):=\lim_{t\to\infty}(f(\mathbf{x}_0+t\boldsymbol{v})-f(\mathbf{x}_0))/t$$

which is well-defined by noting that

$$\begin{aligned}\lim_{t\to\infty}\frac{f(\mathbf{x}_0+t\boldsymbol{v})-f(\mathbf{x}_0)}{t} &= \lim_{t\to\infty}\frac{f(\mathbf{x}_0+t_0\boldsymbol{v})-f(\mathbf{x}_0)}{t}\\ &\quad+\lim_{t\to\infty}\frac{t-t_0}{t}\frac{f(\mathbf{x}_0+t\boldsymbol{v})-f(\mathbf{x}_0+t_0\boldsymbol{v})}{t-t_0},\\ &= \lim_{t\to\infty}\frac{f(\mathbf{x}_0+t\boldsymbol{v})-f(\mathbf{x}_0+t_0\boldsymbol{v})}{t-t_0},\end{aligned}\tag{A34}$$

where $t_0:=\sup_{\{\mathbf{z}\in\mathcal{K}\}}\|\mathbf{z}\|+\|\mathbf{x}_0\|$, and the last term in (A34) is monotone as f coincides with a convex function on $\mathbb{R}^n\setminus\mathcal{K}$. Then, to show that (20) holds, it suffices to verify that $\sup_{\|\boldsymbol{v}\|=1}\phi(\boldsymbol{v})=\mathrm{Lip}(f)=1$. Suppose for contradiction that $\sup_{\|\boldsymbol{v}\|=1}\phi(\boldsymbol{v})\leqslant 1-2\delta$ for some $\delta>0$. Noting that when $t > t_0$, $(f(\mathbf{x}_0 + t\boldsymbol{v}) - f(\mathbf{x}_0 + t_0\boldsymbol{v}))/(t - t_0) \uparrow \phi(\boldsymbol{v})$ as $t \to \infty$, we have for any $t > t_0$ and $\boldsymbol{v}$ with $\|\boldsymbol{v}\| = 1$

$$\begin{aligned}\frac{f(\mathbf{x}_0 + t\boldsymbol{v}) - f(\mathbf{x}_0)}{t} &\leqslant \frac{t_0}{t} + \frac{t - t_0}{t}\frac{f(\mathbf{x}_0 + t\boldsymbol{v}) - f(\mathbf{x}_0 + t_0\boldsymbol{v})}{t - t_0} \\ &\leqslant \frac{t_0}{t} + \frac{t - t_0}{t}\phi(\boldsymbol{v}) < \frac{t_0}{t} + \frac{t - t_0}{t}(1 - 2\delta),\end{aligned} \tag{A35}$$

where the first inequality follows from that f is 1-Lipschitz continuous. Then there exists $t_1 > t_0$ large enough such that $f(\mathbf{x}_0 + t\boldsymbol{v}) - f(\mathbf{x}_0) \leqslant (1 - \delta)t$ for all $t > t_1$ and $\|\boldsymbol{v}\| = 1$. Let

$$B := \max\Big\{0, \sup_{\|\boldsymbol{v}\|=1,\, 0<t\leqslant t_1} \{f(\mathbf{x}_0 + t\boldsymbol{v}) - f(\mathbf{x}_0) - (1 - \delta)t\}\Big\} < \infty.$$

Then, for any $t > 0$ and $\boldsymbol{v} \in \mathbb{R}^n$ with $\|\boldsymbol{v}\| = 1$, we have

$$f(\mathbf{x}_0 + t\boldsymbol{v}) \leqslant f(\mathbf{x}_0) + B + (1 - \delta)t. \tag{A36}$$

Take $F_0 = \delta_{\mathbf{x}_0}$. For $\varepsilon > B/(\delta C_\rho)$ sufficiently large,

$$\begin{aligned}\sup_{F \in \mathbb{B}_1(F_0,\varepsilon)} \rho^F(f(\boldsymbol{\xi})) &= \sup_{\mathbb{E}\|\boldsymbol{\xi} - \mathbf{x}_0\| \leqslant \varepsilon} \rho((f(\boldsymbol{\xi})) \\ &\leqslant \sup_{\mathbb{E}\|\boldsymbol{\xi} - \mathbf{x}_0\| \leqslant \varepsilon} \rho(f(\mathbf{x}_0) + B + (1 - \delta)\|\boldsymbol{\xi} - \mathbf{x}_0\|) \\ &= f(\mathbf{x}_0) + B + (1 - \delta) \sup_{\mathbb{E}\|\boldsymbol{\xi} - \mathbf{x}_0\| \leqslant \varepsilon} \rho(\|\boldsymbol{\xi} - \mathbf{x}_0\|) \\ &< f(\mathbf{x}_0) + C_\rho\varepsilon = \sup_{G \in \mathcal{C}_1(f|F_0,\varepsilon)} \rho^G(X),\end{aligned} \tag{A37}$$

where the first inequality follows from (A36) and the monotonicity of $\rho \in \mathcal{R}_{\text{coh}}$, the second equality follows from the translation invariance and positive homogeneity of $\rho \in \mathcal{R}_{\text{coh}}$, the strict inequality follows from that $\sup_{\mathbb{E}\|\boldsymbol{\xi} - \mathbf{x}_0\| \leqslant \varepsilon} \rho(\|\boldsymbol{\xi} - \mathbf{x}_0\|) = \varepsilon C_\rho$ by the proof of (i) and $\varepsilon > B/(\delta C_\rho)$. Thus, (A37) contradicts (21). Consequently, $\sup_{\|\boldsymbol{v}\|=1} \phi(\boldsymbol{v}) = 1$, i.e., there exists a sequence $\boldsymbol{v}_k$ with $\|\boldsymbol{v}_k\| = 1$ such that $\lim_{k\to\infty} \phi(\boldsymbol{v}_k) = \sup_{\|\boldsymbol{v}\|=1} \phi(\boldsymbol{v}) = 1$. This completes the proof of (ii).

(iii) By (i) and (ii), it suffices to show that if f is convex and Lipschitz continuous, then (20) holds. Note that for any $k \in \mathbb{N}$, there exist $\mathbf{x}_k$ and $\boldsymbol{\beta}_k \in \partial f(\mathbf{x}_k)$ such that $\|\boldsymbol{\beta}_k\|_* > \text{Lip}(f) - 1/k$. Take $\boldsymbol{v}_k$ as a unit vector attaining the dual norm of $\boldsymbol{\beta}_k$, that is, $\|\boldsymbol{v}_k\| = 1$ and $\boldsymbol{v}_k^\top \boldsymbol{\beta}_k = \|\boldsymbol{\beta}_k\|_*$. By the convexity of f, we have for any $\mathbf{x} \in \mathbb{R}^n$

$$\lim_{m\to\infty} \frac{f(\mathbf{x} + m\boldsymbol{v}_k) - f(\mathbf{x})}{m} \geqslant \boldsymbol{\beta}_k^\top \boldsymbol{v}_k = \|\boldsymbol{\beta}_k\|_* > \text{Lip}(f) - \frac{1}{k}, \tag{A38}$$

Therefore, we have (20) holds and complete the proof. $\square$

A.2 Proofs for Section 4

To prove Theorem 5, we need the following lemma from Wu et al. (2022).

Lemma A6. *Let $p \in (1, \infty)$, t, $\varepsilon > 0$, and $\eta \in (0, \varepsilon)$. For $V \in L^p$, the following statements hold.*

(i) If $\|V\|_p \leqslant \varepsilon$, then $\mathbb{E}\left[(|V| + t)^p\right] \leqslant (\varepsilon + t)^p$.

(ii) If $\|V\|_p \leqslant \varepsilon$ and $\mathbb{E}[|V|] \leqslant \varepsilon - \eta$, then there exists $\Delta > 0$ that only depends on p, t, ε, η such that $\mathbb{E}\left[(|V| + t)^p\right] \leqslant (\varepsilon + t)^p - \Delta$. In particular, if p is an integer, then $\mathbb{E}\left[(|V| + t)^p\right] \leqslant (\varepsilon + t)^p - pt^{p-1}\eta$.

Proof of Theorem 5. For $(ii) \Rightarrow (i)$, take $g(\mathbf{x}) = \max_{i \in I}\{c_f \boldsymbol{\beta}_i^\top \mathbf{x} + b_i\}$ and define $\rho(X) = \left(\mathbb{E}[X_+^p]\right)^{1/p}$ which is a monotone risk measure. The result follows immediately from Theorem 3 by noting that

$$\sup_{F \in \mathbb{B}_p(F_0, \varepsilon)} \left(\mathbb{E}^F[f^p(\boldsymbol{\xi})]\right)^{1/p} = \sup_{F \in \mathbb{B}_p(F_0, \varepsilon)} \rho^F(g(\boldsymbol{\xi})).$$

For $(i) \Rightarrow (ii)$, first note that the case $c_f = 0$ is trivial and similar to the case $c_f = 0$ in Theorem 3. It suffices to consider the case $c_f > 0$. We assume without loss of generality that $c_f = 1$. Note that, (27) is equivalent to

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \left(\mathbb{E}\left[f^p\left(\boldsymbol{\xi} + \boldsymbol{\zeta}\right)\right]\right)^{1/p} = \left(\mathbb{E}^{F_0}\left[f^p\left(\boldsymbol{\zeta}\right)\right]\right)^{1/p} + \varepsilon, \ \forall \boldsymbol{\zeta} \sim F_0, \varepsilon > 0. \tag{A39}$$

For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, take $\boldsymbol{\zeta} \sim F_0 := \delta_{\mathbf{x}}$ and set $\varepsilon := \|\mathbf{y} - \mathbf{x}\|$. Then, by (A39),

$$f(\mathbf{y}) \leqslant \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \left(\mathbb{E}\left[f^p(\mathbf{x} + \boldsymbol{\xi})\right]\right)^{1/p} = f(\mathbf{x}) + \varepsilon = f(\mathbf{x}) + \|\mathbf{y} - \mathbf{x}\|.$$

Exchanging $\mathbf{x}$ and $\mathbf{y}$ yields $|f(\mathbf{y}) - f(\mathbf{x})| \leqslant \|\mathbf{y} - \mathbf{x}\|$, and hence $\mathrm{Lip}(f) \leqslant 1$. We then aim to show the following statement by contradiction:

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) = \varepsilon, \qquad \forall \mathbf{x} \in \{\mathbf{x} : f(\mathbf{x}) > 0\}, \ \forall \varepsilon > 0. \tag{A40}$$

Suppose to the contrary that there exist $\mathbf{x}_0 \in \mathbb{R}^n$ with $f(\mathbf{x}_0) > 0$ and $\varepsilon_0 > 0$ such that

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0) < \varepsilon_0. \tag{A41}$$

For $\varepsilon > 0$, define

$$k_\varepsilon = \sup_{\|\mathbf{y}\| \leqslant 2\varepsilon} \left\{ \frac{f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0)}{2\varepsilon} \right\}.$$

Note that for any $\mathbf{z} \in \mathbb{R}^n$, by the convexity of f, we have $(f(\mathbf{x}_0 + t\mathbf{z}) - f(\mathbf{x}_0))/t$ is increasing in $t \in \mathbb{R}_+$, and thus, for any $\varepsilon \in (0, \varepsilon_0/2)$, it holds that

$$f(\mathbf{x}_0 + \mathbf{z}) - f(\mathbf{x}_0) \leqslant \frac{2\varepsilon}{\varepsilon_0} \left(f\left(\mathbf{x}_0 + \frac{\varepsilon_0}{2\varepsilon}\mathbf{z} \right) - f(\mathbf{x}_0) \right).$$

Taking supremum over $\|\mathbf{z}\| \leqslant 2\varepsilon$ to both sides yields that for $\varepsilon \in (0, \varepsilon_0/2)$,

$$k_\varepsilon \leqslant \sup_{\|\mathbf{z}\| \leqslant 2\varepsilon} \frac{f\left(\mathbf{x}_0 + \frac{\varepsilon_0}{2\varepsilon}\mathbf{z}\right) - f(\mathbf{x}_0)}{\varepsilon_0} \leqslant \sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} \left\{ \frac{f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0)}{\varepsilon_0} \right\} < 1, \tag{A42}$$

where the second inequality follows from $\|\varepsilon_0 \mathbf{z}/(2\varepsilon)\| \leqslant \varepsilon_0$ whenever $\|\mathbf{z}\| \leqslant 2\varepsilon$. Similarly, for any $\mathbf{y} \in \mathbb{R}^n$ with $\|\mathbf{y}\| \leqslant 2\varepsilon$, we have

$$f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}) \leqslant \frac{\|\mathbf{y}\|}{2\varepsilon} \left(f\left(\mathbf{x}_0 + \frac{2\varepsilon}{\|\mathbf{y}\|}\mathbf{y} \right) - f(\mathbf{x}) \right),$$

which implies that

$$\sup_{\|\mathbf{y}\| \leqslant 2\varepsilon} \left\{ \frac{f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0)}{\|\mathbf{y}\|} \right\} \leqslant \sup_{\|\mathbf{y}\| \leqslant 2\varepsilon} \left\{ \frac{f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0)}{2\varepsilon} \right\} = k_\varepsilon. \tag{A43}$$

Thus, by (A43), we have for $\varepsilon \in (0, \varepsilon_0/2)$,

$$f(\mathbf{x}_0 + \mathbf{y}) - f(\mathbf{x}_0) \leqslant k_\varepsilon \|\mathbf{y}\|, \quad \forall \, \|\mathbf{y}\| \leqslant 2\varepsilon. \tag{A44}$$

Further, note that

$$(f(\mathbf{x}_0) + \|\mathbf{y}\|)^p - (f(\mathbf{x}_0) + k_\varepsilon \|\mathbf{y}\|)^p \geqslant p f^{p-1}(\mathbf{x}_0)(1 - k_\varepsilon)\|\mathbf{y}\|, \, \forall \, \mathbf{y} \in \mathbb{R}^n. \tag{A45}$$

Therefore, for $\varepsilon \in (0, \varepsilon_0/2)$,

$$\begin{aligned}
&\sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \mathbb{E}\left[f^p\left(\mathbf{x}_0+\boldsymbol{\xi}\right)\right] \\
&= \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \left\{\mathbb{E}\left[f^p\left(\mathbf{x}_0+\boldsymbol{\xi}\right) \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant 2\varepsilon\}}\right]+\mathbb{E}\left[f^p\left(\mathbf{x}_0+\boldsymbol{\xi}\right) \mathbb{1}_{\{\|\boldsymbol{\xi}\|>2\varepsilon\}}\right]\right\} \\
&\leqslant \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \left\{\mathbb{E}\left[\left(f\left(\mathbf{x}_0\right)+k_{\varepsilon}\|\boldsymbol{\xi}\|\right)^p \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant 2\varepsilon\}}\right]+\mathbb{E}\left[f^p\left(\mathbf{x}_0+\boldsymbol{\xi}\right) \mathbb{1}_{\{\|\boldsymbol{\xi}\|>2\varepsilon\}}\right]\right\} \\
&\leqslant \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \left\{\mathbb{E}\left[\left(f\left(\mathbf{x}_0\right)+k_{\varepsilon}\|\boldsymbol{\xi}\|\right)^p \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant 2\varepsilon\}}\right]+\mathbb{E}\left[\left(f\left(\mathbf{x}_0\right)+\|\boldsymbol{\xi}\|\right)^p \mathbb{1}_{\{\|\boldsymbol{\xi}\|>2\varepsilon\}}\right]\right\} \\
&= \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \left\{\mathbb{E}\left[\left(f\left(\mathbf{x}_0\right)+\|\boldsymbol{\xi}\|\right)^p\right]-\mathbb{E}\left[\left(\left(f\left(\mathbf{x}_0\right)+\|\boldsymbol{\xi}\|\right)^p-\left(f\left(\mathbf{x}_0\right)+k_{\varepsilon}\|\boldsymbol{\xi}\|\right)^p\right) \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant 2\varepsilon\}}\right]\right\} \\
&\leqslant \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \left\{\mathbb{E}\left[\left(f\left(\mathbf{x}_0\right)+\|\boldsymbol{\xi}\|\right)^p\right]-p f^{p-1}\left(\mathbf{x}_0\right)\left(1-k_{\varepsilon}\right) \mathbb{E}\left[\|\boldsymbol{\xi}\| \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant 2\varepsilon\}}\right]\right\} \\
&= \sup_{\mathbb{E}[|V|^p] \leqslant \varepsilon^p} \left\{\mathbb{E}\left[\left(f\left(\mathbf{x}_0\right)+|V|\right)^p\right]-p f^{p-1}\left(\mathbf{x}_0\right)\left(1-k_{\varepsilon}\right) \mathbb{E}\left[|V| \mathbb{1}_{\{|V| \leqslant 2\varepsilon\}}\right]\right\} \\
&=: I,
\end{aligned}$$

where the first inequality follows from (A44), the second inequality holds because f is nonnegative with $\operatorname{Lip}(f) \leqslant 1$; the third inequality follows from (A45), and the third equality holds since the objective function in the optimization problem depends only on $\|\boldsymbol{\xi}\|$. Define

$$\begin{aligned}
&\mathcal{V}_1=\left\{V \in L^p: \mathbb{E}|V|^p \leqslant \varepsilon^p,\, \mathbb{E}[|V| \mathbb{1}_{\{|V| \leqslant 2\varepsilon\}}] \leqslant \frac{(1-2^{-p/q})}{2} \varepsilon\right\}, \\
&\mathcal{V}_2=\left\{V \in L^p: \mathbb{E}|V|^p \leqslant \varepsilon^p\right\} \setminus \mathcal{V}_1 .
\end{aligned}$$

We can rewrite $I=\max\{I_1, I_2\}$ with

$$I_i=\sup_{V \in \mathcal{V}_i}\left\{\mathbb{E}\left[\left(f\left(\mathbf{x}_0\right)+|V|\right)^p\right]-p f^{p-1}\left(\mathbf{x}_0\right)\left(1-k_{\varepsilon}\right) \mathbb{E}\left[|V| \mathbb{1}_{\{|V| \leqslant 2\varepsilon\}}\right]\right\}, \quad i \in [2].$$

One can verify that for any $V \in \mathcal{V}_1$,

$$\begin{aligned}
\mathbb{E}[|V|] &= \mathbb{E}[|V| \mathbb{1}_{\{|V|>2\varepsilon\}}]+\mathbb{E}[|V| \mathbb{1}_{\{|V| \leqslant 2\varepsilon\}}] \\
&\leqslant \varepsilon 2^{-p/q}+\frac{\left(1-2^{-p/q}\right) \varepsilon}{2}=\frac{\left(1+2^{-p/q}\right) \varepsilon}{2}<\varepsilon,
\end{aligned} \tag{A46}$$

where the first inequality follows from the Hölder's inequality, Markov's inequality, and the definition of $\mathcal{V}_1$. It holds that

$$I_1 \leqslant \sup_{V\in\mathcal{V}_1} \mathbb{E}\left[(f(\mathbf{x}_0)+|V|)^p\right] < (f(\mathbf{x}_0)+\varepsilon)^p, \tag{A47}$$

where the strict inequality follows from (A46) and Statement (ii) of Lemma A6 by noting that $f(\mathbf{x}_0)>0$. For I_2, we have

$$\begin{aligned}
I_2 &\leqslant \sup_{V\in\mathcal{V}_2} \mathbb{E}\left[(f(\mathbf{x}_0)+|V|)^p\right] - \inf_{V\in\mathcal{V}_2} pf^{p-1}(\mathbf{x}_0)(1-k_\varepsilon)\mathbb{E}\left[|V|\mathbb{1}_{\{|V|\leqslant 2\varepsilon\}}\right] \\
&\leqslant (f(\mathbf{x}_0)+\varepsilon)^p - pf^{p-1}(\mathbf{x}_0)(1-k_\varepsilon)\inf_{V\in\mathcal{V}_2}\mathbb{E}\left[|V|\mathbb{1}_{\{|V|\leqslant 2\varepsilon\}}\right] \\
&\leqslant (f(\mathbf{x}_0)+\varepsilon)^p - pf^{p-1}(\mathbf{x}_0)(1-k_\varepsilon)\frac{\left(1-2^{-p/q}\right)\varepsilon}{2} \\
&< (f(\mathbf{x}_0)+\varepsilon)^p,
\end{aligned} \tag{A48}$$

where the second inequality follows from Statement (i) of Lemma A6, and the third inequality is due to the definition of $\mathcal{V}_2$. Combining (A47) and (A48), we have

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon} \mathbb{E}[f^p(\mathbf{x}_0+\boldsymbol{\xi})] \leqslant I = \max\{I_1, I_2\} < (f(\mathbf{x}_0)+\varepsilon)^p,$$

which yields a contradiction to (A39). Hence, (A40) holds. By the arguments to prove (A18) as in the proof of Proposition 3, for any $\mathbf{z}\in\mathbb{R}^n$ with $f(\mathbf{z})>0$, there exists a subgradient $\boldsymbol{\eta}_{\mathbf{z}}\in\partial f(\mathbf{z})$ such that $\|\boldsymbol{\eta}_{\mathbf{z}}\|_*=1$. Moreover, for any $\mathbf{z}\in\mathbb{R}^n$ with $f(\mathbf{z})=0$, we take $\boldsymbol{\eta}_{\mathbf{z}}=\mathbf{0}\in\partial f(\mathbf{z})$. By the convexity of f, for any $\mathbf{x}\in\mathbb{R}^n$,

$$\begin{aligned}
f(\mathbf{x}) &= \max_{\mathbf{z}\in\mathbb{R}^n}\{f(\mathbf{z})+\boldsymbol{\eta}_{\mathbf{z}}^\top(\mathbf{x}-\mathbf{z})\} \\
&= \max\Big\{\sup_{f(\mathbf{z})=0}\{f(\mathbf{z})+\boldsymbol{\eta}_{\mathbf{z}}^\top(\mathbf{x}-\mathbf{z})\},\ \sup_{f(\mathbf{z})>0}\{f(\mathbf{z})+\boldsymbol{\eta}_{\mathbf{z}}^\top(\mathbf{x}-\mathbf{z})\}\Big\} \\
&= \max\Big\{0,\ \sup_{f(\mathbf{z})>0}\{f(\mathbf{z})+\boldsymbol{\eta}_{\mathbf{z}}^\top(\mathbf{x}-\mathbf{z})\}\Big\} =: (s(\mathbf{x}))_+,
\end{aligned}$$

where $s(\mathbf{x}) := \sup_{(\boldsymbol{\beta},b)\in\mathcal{I}}\{\boldsymbol{\beta}^\top\mathbf{x}+b\}$ with $\mathcal{I} := \{(\boldsymbol{\eta}_{\mathbf{z}},\ f(\mathbf{z})-\boldsymbol{\eta}_{\mathbf{z}}^\top\mathbf{z}) : f(\mathbf{z})>0\}$. Fix $\mathbf{x}$ with $f(\mathbf{x})=0$. Take $\{(\boldsymbol{\eta}^{\mathbf{x}}_{\mathbf{z}_k}, b^{\mathbf{x}}_{\mathbf{z}_k})\}_{k\in\mathbb{N}}\in\mathcal{I}$ such that ${\boldsymbol{\eta}^{\mathbf{x}}_{\mathbf{z}_k}}^\top\mathbf{x}+b^{\mathbf{x}}_{\mathbf{z}_k}\to s(\mathbf{x})$. Note that $\|\boldsymbol{\eta}^{\mathbf{x}}_{\mathbf{z}_k}\|_*=1$ and ${\boldsymbol{\eta}^{\mathbf{x}}_{\mathbf{z}_k}}^\top\mathbf{y}+b^{\mathbf{x}}_{\mathbf{z}_k}\leqslant f(\mathbf{y})$ for all $\mathbf{y}\in\mathbb{R}^n$ and $k\in\mathbb{N}$. By Hölder's inequality, $|{\boldsymbol{\eta}^{\mathbf{x}}_{\mathbf{z}_k}}^\top\mathbf{x}|\leqslant\|\mathbf{x}\|$, and hence the sequence $b^{\mathbf{x}}_{\mathbf{z}_k}$ is bounded. Thus, by the Bolzano–Weierstrass theorem, up to a subsequence, $(\boldsymbol{\eta}^{\mathbf{x}}_{\mathbf{z}_k}, b^{\mathbf{x}}_{\mathbf{z}_k})\to(\bar{\boldsymbol{\eta}}^{\mathbf{x}},\bar{b}^{\mathbf{x}})$. By the continuity of $(\boldsymbol{\eta},b)\mapsto\boldsymbol{\eta}^\top\mathbf{y}+b$, for any fixed $\mathbf{y}\in\mathbb{R}^n$, we have ${\bar{\boldsymbol{\eta}}^{\mathbf{x}}}^\top\mathbf{y}+\bar{b}^{\mathbf{x}}=\lim_{n\to\infty}{\boldsymbol{\eta}^{\mathbf{x}}_{\mathbf{z}_k}}^\top\mathbf{y}+b^{\mathbf{x}}_{\mathbf{z}_k}\leqslant s(\mathbf{y})\leqslant f(\mathbf{y})$, and in particular ${\bar{\boldsymbol{\eta}}^{\mathbf{x}}}^\top\mathbf{x}+\bar{b}^{\mathbf{x}}=s(\mathbf{x})\leqslant 0$. Define $\widetilde{\mathcal{I}}=\mathcal{I}\cup\{(\bar{\boldsymbol{\eta}}^{\mathbf{x}},\bar{b}^{\mathbf{x}}) : \mathbf{x}\in\{\mathbf{x} : f(\mathbf{x})=0\}\}$. It is straightforward to verify that $s(\mathbf{x})=\max_{(\boldsymbol{\beta},b)\in\widetilde{\mathcal{I}}}\{\boldsymbol{\beta}^\top\mathbf{x}+b\}$ for any $\mathbf{x}\in\mathbb{R}^n$. Then, we complete the proof. $\square$

To prove Proposition 6, we need the following three lemmas.

Lemma A7. *For $p > 1$ and $c > 1$, let $\mathcal{H}_p$ be defined in (29) with $\ell(z, t) = c(z - t)_+$, and X be a random variable with $\mathbb{P}(X = a_1) = 1 - \mathbb{P}(X = a_0) = \pi$, $a_1 > a_0$. If $\pi < c^{-p}$, then problem $\inf_{t\in\mathbb{R}}\{t + c(\mathbb{E}[(X - t)_+^p])^{1/p}\}$ admits a unique minimizer t^* on $(-\infty, a_0)$.*

Proof. Define $H(t) := t + c(\mathbb{E}[(X - t)_+^p])^{1/p}$, $t \in \mathbb{R}$. Note that

$$H(t) = \begin{cases} t + c\Big((1-\pi)(a_0 - t)^p + \pi(a_1 - t)^p\Big)^{1/p}, & t < a_0, \\ (1 - c\pi^{1/p})t + c\pi^{1/p}a_1, & a_0 \leqslant t < a_1, \\ t, & t \geqslant a_1. \end{cases}$$

By $\pi < c^{-p}$, we have H is continuous and strictly increasing on $[a_0, \infty)$. For $t < a_0$,

$$H'(t) = 1 - c((1-\pi)(a_0 - t)^p + \pi(a_1 - t)^p)^{1/p-1}((1-\pi)(a_0 - t)^{p-1} + \pi(a_1 - t)^{p-1}).$$

This implies $\lim_{t\uparrow a_0} H'(t) = 1 - c\pi^{1/p} > 0$, and hence no minimizer lies in $[a_0, \infty)$. Also, note that $\big(\mathbb{E}[(X - t)_+^p]\big)^{1/p}$ is strictly convex in $t \in (-\infty, a_0]$, and thus, H is strictly convex on $(-\infty, a_0]$. By $\lim_{t\to-\infty} H(t) = \infty$, we have H admits a unique minimizer $t^* \in (-\infty, a_0)$. This completes the proof. $\square$

Lemma A8. *For $p \in (1, \infty)$, $a > 0$, $k \in [0, 1)$, $\varepsilon > 0$, and a random variable $V \geqslant 0$, there exists $c \in (0, \infty)$ depending only on (p, a, ε) such that*

$$I := (\mathbb{E}[(a + kV\mathbb{1}_{\{V\leqslant\varepsilon\}} + V\mathbb{1}_{\{V>\varepsilon\}})^p])^{1/p} \leqslant a + k(\mathbb{E}[V^p])^{1/p} + c\mathbb{E}[V^p].$$

Proof. Denote by $v = (\mathbb{E}[V^p])^{1/p}$. Note that

$$\begin{aligned} I^p &= \mathbb{E}[(a + kV)^p\mathbb{1}_{\{V\leqslant\varepsilon\}}] + \mathbb{E}[(a + V)^p\mathbb{1}_{\{V>\varepsilon\}}] \\ &\leqslant \mathbb{E}[(a + kV)^p] + \mathbb{E}[(a + V)^p\mathbb{1}_{\{V>\varepsilon\}}] \\ &\leqslant \Big(a + k(\mathbb{E}[V^p])^{1/p}\Big)^p + \mathbb{E}[(a + V)^p\mathbb{1}_{\{V>\varepsilon\}}] \\ &= (a + kv)^p + \mathbb{E}\left[\left(\frac{a + V}{V}\right)^p V^p\mathbb{1}_{\{V>\varepsilon\}}\right] \\ &\leqslant (a + kv)^p + \left(\frac{a + \varepsilon}{\varepsilon}\right)^p \mathbb{E}[V^p\mathbb{1}_{\{V>\varepsilon\}}] \leqslant (a + kv)^p + c_1v^p, \end{aligned} \tag{A49}$$

where the second inequality follows from Lemma A6(i) (Jensen's inequality), the third inequality follows from that $(a+x)/x$ is decreasing in $x \in [\varepsilon, \infty)$, and $c_1 := (a+\varepsilon)^p/\varepsilon^p$. It then follows that

$$I \leqslant \left([a+kv]^p + c_1 v^p\right)^{1/p} \leqslant a + kv + \frac{c_1 v^p}{p\,(a+kv)^{p-1}} \leqslant a + kv + \frac{c_1 v^p}{pa^{p-1}},$$

where the second inequality follows from $(b^p + x)^{1/p} \leqslant b + x/(pb^{p-1})$ by the concavity of $(b^p+x)^{1/p}$ on $x \in \mathbb{R}_+$ for $b > 0$, and the last inequality follows from $kv_p \geqslant 0$. Setting $c = c_1/(pa^{p-1})$ completes the proof. □

Lemma A9. *For $p > 1$, $a_0 > 0$, $a_1 > 0$, $\bar{\varepsilon} \in (0,1]$, $\pi \in (0,1)$, $k \in [0,1)$, and $c > 0$, let $\Phi : \mathbb{R}_+ \times K \to \mathbb{R}$ be given by*

$$\Phi(\varepsilon; \boldsymbol{u}) := \left((1-\pi)\left(a_0 + k\varepsilon u_0 + c\varepsilon^p u_0^p\right)^p + \pi(a_1 + \varepsilon u_1)^p\right)^{1/p},$$

where $K := \{\boldsymbol{u} \in \mathbb{R}_+^2 : (1-\pi)u_0^p + \pi u_1^p \leqslant 1\}$. Then there exist $L < 1$ and $M < \infty$ depending only on $(p, a_0, a_1, \bar{\varepsilon}, \pi, k_0, c)$ such that

$$\sup_{\boldsymbol{u} \in K} \Phi(\varepsilon; \boldsymbol{u}) \leqslant \left((1-\pi)a_0^p + \pi a_1^p\right)^{1/p} + L\varepsilon + M(\varepsilon^2 + \varepsilon^p), \quad \forall\, \varepsilon \in [0, \bar{\varepsilon}].$$

Proof. Define $\Phi_0(\varepsilon; \boldsymbol{u}) := \left((1-\pi)(a_0 + k\varepsilon u_0)^p + \pi(a_1 + \varepsilon u_1)^p\right)^{1/p}$. By Minkowski's inequality,

$$\Phi(\varepsilon; \boldsymbol{u}) \leqslant \Phi_0(\varepsilon; \boldsymbol{u}) + c\varepsilon^p u_0^p. \tag{A50}$$

Thus, it suffices to show that there exist $L < 1$ and $M_1 < \infty$ such that

$$\sup_{\boldsymbol{u} \in K} \Phi_0(\varepsilon; \boldsymbol{u}) \leqslant \left((1-\pi)a_0^p + \pi a_1^p\right)^{1/p} + L\varepsilon + M_1\varepsilon^2, \quad \forall\, \varepsilon \in [0, \bar{\varepsilon}]. \tag{A51}$$

To show that (A51) holds, we consider a second-order Taylor expansion of $\Phi_0(\varepsilon; \boldsymbol{u})$ at $\varepsilon = 0$. For any $\boldsymbol{u} \in K$, there exists $z(\varepsilon, \boldsymbol{u}) \in [0, \varepsilon]$ such that

$$\Phi_0(\varepsilon; \boldsymbol{u}) = \Phi_0(0; \boldsymbol{u}) + \varepsilon\partial_\varepsilon\Phi_0(0; \boldsymbol{u}) + \partial_\varepsilon^2\Phi_0(z(\varepsilon, \boldsymbol{u}); \boldsymbol{u})\frac{\varepsilon^2}{2},$$

where $\partial_\varepsilon\Phi_0$ and $\partial_\varepsilon^2\Phi_0$ denote the first- and second-order partial derivatives of Φ_0 with respect to ε, respectively. Note that $\Phi_0(0; \boldsymbol{u}) = \left((1-\pi)a_0^p + \pi a_1^p\right)^{1/p}$ which is independent of $\boldsymbol{u}$. Taking the supremum over $\boldsymbol{u} \in K$ yields, for any $\varepsilon \in [0, \bar{\varepsilon}]$,

$$\begin{aligned}
\sup_{\boldsymbol{u}\in K} \Phi(\varepsilon; \boldsymbol{u}) &\leqslant \Phi_0(0; \boldsymbol{u}) + \varepsilon \sup_{\boldsymbol{u}\in K} \{\partial_\varepsilon \Phi_0(0; \boldsymbol{u})\} + \sup_{\boldsymbol{u}\in K} \left\{\partial_\varepsilon^2 \Phi_0(z(\varepsilon, \boldsymbol{u}); \boldsymbol{u})\right\} \frac{\varepsilon^2}{2} \\
&\leqslant \Phi_0(0; \boldsymbol{u}) + \varepsilon \sup_{\boldsymbol{u}\in K} \{\partial_\varepsilon \Phi_0(0; \boldsymbol{u})\} + \sup_{\boldsymbol{u}\in K, s\in[0,\bar{\varepsilon}]} \left\{|\partial_\varepsilon^2 \Phi_0(s; \boldsymbol{u})|\right\} \frac{\varepsilon^2}{2} \\
&= \Phi_0(0; \boldsymbol{u}) + L\varepsilon + M_2 \frac{\varepsilon^2}{2},
\end{aligned}$$

where the second inequality uses the fact that $z(\varepsilon, \boldsymbol{u}) \in [0, \bar{\varepsilon}]$, and we define $L := \sup_{\boldsymbol{u}\in K} \{\partial_\varepsilon \Phi_0(0; \boldsymbol{u})\}$ and $M_2 := \sup_{(s,u)\in[0,\bar{\varepsilon}]\times K} |\partial_\varepsilon^2 \Phi_0(s; u)|$. Since $\partial_\varepsilon^2 \Phi_0$ is continuous on the compact set $[0, \bar{\varepsilon}] \times K$, it follows that $M_2 < \infty$. Hence, it remains to show that $L < 1$. A direct computation gives

$$\begin{aligned}
L &= \sup_{\boldsymbol{u}\in K} \left\{ \frac{(1-\pi)a_0^{p-1}k_0 u_0 + \pi a_1^{p-1} u_1}{((1-\pi)a_0^p + \pi a_1^p)^{(p-1)/p}} \right\} \\
&\leqslant \sup_{\boldsymbol{u}\in K} \left\{ \frac{\left((1-\pi)(a_0^{p-1}k_0)^q + \pi a_1^{pq-q}\right)^{1/q} ((1-\pi)u_0^p + \pi u_1^p)^{1/p}}{((1-\pi)a_0^p + \pi a_1^p)^{(p-1)/p}} \right\} \\
&\leqslant \frac{((1-\pi)a_0^p k_0^q + \pi a_1^p)^{1/q}}{((1-\pi)a_0^p + \pi a_1^p)^{1/q}} < 1,
\end{aligned}$$

where the first inequality follows from Hölder's inequality with $q := p/(p-1)$, the second inequality uses the constraint $(1-\pi)u_0^p + \pi u_1^p \leqslant 1$ for $\boldsymbol{u} \in K$, and the strict inequality holds since $k_0 < 1$. Setting $M_1 := M_2/2$, (A51) follows. By (A50) and (A51), for any $\varepsilon \in [0, \bar{\varepsilon}]$,

$$\begin{aligned}
\sup_{\boldsymbol{u}\in K} \Phi(\varepsilon; \boldsymbol{u}) &\leqslant \sup_{\boldsymbol{u}\in K} \{\Phi_0(\varepsilon; \boldsymbol{u})\} + c\varepsilon^p \sup_{\boldsymbol{u}\in K} u_0^p \\
&\leqslant \left((1-\pi)a_0^p + \pi a_1^p\right)^{1/p} + L\varepsilon + M(\varepsilon^2 + \varepsilon^p),
\end{aligned}$$

where the second inequality follows from $\sup_{\boldsymbol{u}\in K} u_0^p \leqslant (1-\pi)^{-1}$ and $M := \max\{M_1, c(1-\pi)^{-1}\}$. This completes the proof. □

Proof of Proposition 6. For the "if" part, the result follows directly from Corollary 1.

For the "only if" part, we first consider the case $c_f = 0$. In this case, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, take $\boldsymbol{\zeta} \sim F_0 = \delta_{\mathbf{x}}$ and $\varepsilon := \|\mathbf{x} - \mathbf{y}\|$. By (31), we have

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \mathcal{H}_p(f(\boldsymbol{\xi} + \mathbf{x})) = f(\mathbf{x}).$$

It follows that $f(\mathbf{y}) \leqslant f(\mathbf{x})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, and thus, f is a constant function, which completes the proof for $c_f = 0$. We now turn to the case $c_f > 0$. Without loss of generality, we assume $c_f = 1$. In this case, (31) is equivalent to

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon^p} \mathcal{H}_p(f(\boldsymbol{\xi}+\boldsymbol{\zeta})) = \mathcal{H}_p(f(\boldsymbol{\zeta})) + c\varepsilon, \qquad \forall\, \boldsymbol{\zeta} \sim F_0,\ \varepsilon > 0. \tag{A52}$$

We next show that for any $\mathbf{x} \in \mathbb{R}^n$, the subgradient $\nabla f(\mathbf{x}) \in \partial f(\mathbf{x})$ satisfies $\|\nabla f(\mathbf{x})\|_* \leqslant 1$, where $\partial f(\mathbf{x})$ denotes the subdifferential of f at $\mathbf{x}$. Suppose, for contradiction, that there exists $\mathbf{x}_0 \in \mathbb{R}^n$ such that $\|\nabla f(\mathbf{x}_0)\|_* > 1$. By definition of the dual norm, there exists $\boldsymbol{\eta}_{\mathbf{x}_0} \in \mathbb{R}^n$ with $\|\boldsymbol{\eta}_{\mathbf{x}_0}\| = 1$ such that $\boldsymbol{\eta}_{\mathbf{x}_0}^\top \nabla f(\mathbf{x}_0) = \|\nabla f(\mathbf{x}_0)\|_*$. For $\varepsilon > 0$, define $\boldsymbol{\xi}_0 \sim \pi\delta_{c\varepsilon\boldsymbol{\eta}_{\mathbf{x}_0}} + (1-\pi)\delta_0$ with $\pi = c^{-p}$. Then $\mathbb{E}[\|\boldsymbol{\xi}_0\|^p] = \varepsilon^p$, and thus,

$$\begin{aligned}
&\sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon^p} \mathcal{H}_p(f(\boldsymbol{\xi}+\mathbf{x}_0)) \\
&\geqslant \mathcal{H}_p(f(\boldsymbol{\xi}_0+\mathbf{x}_0)) \\
&= f(\mathbf{x}_0) + \mathcal{H}_p(f(\boldsymbol{\xi}_0+\mathbf{x}_0) - f(\mathbf{x}_0)) \\
&= f(\mathbf{x}_0) + \inf_t \{t + c(\mathbb{E}[(f(\boldsymbol{\xi}_0+\mathbf{x}_0) - f(\mathbf{x}_0) - t)_+^p])^{1/p}\} \\
&= f(\mathbf{x}_0) + \inf_t \{t + c(\mathbb{E}[((f(\boldsymbol{\xi}_0+\mathbf{x}_0) - f(\mathbf{x}_0))\mathbb{1}_{\{\boldsymbol{\xi}_0 = c\varepsilon\boldsymbol{\eta}_{\mathbf{x}_0}\}} - t)_+^p])^{1/p}\} \\
&\geqslant f(\mathbf{x}_0) + \inf_t \{t + c(\mathbb{E}[(c\varepsilon\boldsymbol{\eta}_{\mathbf{x}_0}^\top \nabla f(\mathbf{x}_0)\mathbb{1}_{\{\boldsymbol{\xi}_0 = c\varepsilon\boldsymbol{\eta}_{\mathbf{x}_0}\}} - t)_+^p])^{1/p}\} \\
&= f(\mathbf{x}_0) + c\varepsilon\|\nabla f(\mathbf{x}_0)\|_* \inf_t \{t + c(\mathbb{E}[(\mathbb{1}_{\{\boldsymbol{\xi}_0 = c\varepsilon\boldsymbol{\eta}_{\mathbf{x}_0}\}} - t)_+^p])^{1/p}\} \\
&= f(\mathbf{x}_0) + c\varepsilon\|\nabla f(\mathbf{x}_0)\|_* > f(\mathbf{x}_0) + c\varepsilon,
\end{aligned} \tag{A53}$$

where the first equality follows from the translation invariance of $\mathcal{H}_p$, the second inequality follows from the convexity of f, the forth equality follows from $\boldsymbol{\eta}_{\mathbf{x}_0}^\top \nabla f(\mathbf{x}_0) = \|\nabla f(\mathbf{x}_0)\|_*$ and the positive homogeneity of $\mathcal{H}_p$, the fifth equality uses $\mathcal{H}_p(\mathbb{1}_{\{\boldsymbol{\xi}_0 = c\varepsilon\boldsymbol{\eta}_{\mathbf{x}_0}\}}) = 1$, and the strict inequality follows from $\|\nabla f(\mathbf{x}_0)\|_* > 1$ and $\varepsilon > 0$. The inequality (A53) contradicts (A52). Therefore, $\|\nabla f(\mathbf{x})\|_* \leqslant 1$ for all $\mathbf{x} \in \mathbb{R}^n$. Since f is convex, it follows that f is Lipschitz continuous with $\mathrm{Lip}(f) \leqslant 1$. Finally, by the proof of Proposition 3, to establish that f satisfies (12), it suffices to show that

$$\sup_{\|\mathbf{y}\|\leqslant\varepsilon} f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) = \varepsilon, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\ \varepsilon > 0.$$

Suppose, to the contrary, that there exist $\mathbf{x}_0 \in \mathbb{R}^n$ and $\varepsilon_0 > 0$ such that

$$\sup_{\|\mathbf{y}\|\leqslant\varepsilon_0} f(\mathbf{x}_0+\mathbf{y}) - f(\mathbf{x}_0) < \varepsilon_0. \tag{A54}$$

Define

$$k_0 := \sup_{\|\mathbf{y}\|\leqslant\varepsilon_0} \frac{f(\mathbf{x}_0+\mathbf{y})-f(\mathbf{x}_0)}{\varepsilon_0} < 1.$$

By the convexity of f, we have

$$f(\mathbf{x}_0+\mathbf{y})-f(\mathbf{x}_0) \leqslant k_0\|\mathbf{y}\|, \qquad \forall\, \|\mathbf{y}\| \leqslant \varepsilon_0. \tag{A55}$$

By (A52), there must exist $\mathbf{x}_1 \in \mathbb{R}^n$ such that $f(\mathbf{x}_1) > f(\mathbf{x}_0)$. Let $\boldsymbol{\zeta} \sim F_0 := (1-\pi)\delta_{\mathbf{x}_0} + \pi\delta_{\mathbf{x}_1}$, with some $\pi \in (0, c^{-p})$. By Lemma A7, there exists a unique minimizer $t^* < f(\mathbf{x}_0)$ such that

$$\mathcal{H}_p(f(\boldsymbol{\zeta})) = t^* + c(\mathbb{E}[(f(\boldsymbol{\zeta})-t^*)^p])^{1/p}. \tag{A56}$$

Fix $\bar{\varepsilon} := \min\{1, \varepsilon_0\}$. For t^* satisfying (A56) and for all sufficiently small $\varepsilon < \bar{\varepsilon}$, we have

$$\begin{aligned}
&\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p} \mathcal{H}_p(f(\boldsymbol{\zeta}+\boldsymbol{\xi})) \\
&= \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon^p} \inf_t \{t + c(\mathbb{E}[(f(\boldsymbol{\xi}+\boldsymbol{\zeta})-t)_+^p])^{1/p}\} \\
&\leqslant \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon^p} \{t^* + c(\mathbb{E}[(f(\boldsymbol{\xi}+\boldsymbol{\zeta})-t^*)_+^p])^{1/p}\} \\
&= \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon^p} \{t^* + c(\mathbb{E}[(f(\boldsymbol{\xi}+\boldsymbol{\zeta})-f(\boldsymbol{\zeta})+f(\boldsymbol{\zeta})-t^*)_+^p])^{1/p}\} \\
&\leqslant \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon^p} \big\{t^* + c\big(\mathbb{E}[\big((f(\boldsymbol{\xi}+\boldsymbol{\zeta})-f(\boldsymbol{\zeta}))_+ + (f(\boldsymbol{\zeta})-t^*)_+\big)^p]\big)^{1/p}\big\} \\
&= \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p} \big\{t^* + c\big(\mathbb{E}[(T(\boldsymbol{\xi},\boldsymbol{\zeta}) + (f(\boldsymbol{\zeta})-t^*)_+)^p]\big)^{1/p}\big\},
\end{aligned} \tag{A57}$$

where the last inequality uses $(x+y)_+ \leqslant x_+ + y_+$ for all $x, y \in \mathbb{R}$, and $T(\boldsymbol{\xi},\boldsymbol{\zeta}) := (f(\boldsymbol{\xi}+\boldsymbol{\zeta})-f(\boldsymbol{\zeta}))_+$. It therefore suffices to consider

$$I := \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p]\leqslant\varepsilon^p} (\mathbb{E}[((f(\boldsymbol{\zeta})-t^*)_+ + T(\boldsymbol{\xi},\boldsymbol{\zeta}))^p])^{1/p}.$$

Note that, by (A55), we have $T(\boldsymbol{\xi},\mathbf{x}_0) \leqslant k_0\|\boldsymbol{\xi}\|\mathbb{1}_{\{\|\boldsymbol{\xi}\|\leqslant\varepsilon_0\}} + \|\boldsymbol{\xi}\|\mathbb{1}_{\{\|\boldsymbol{\xi}\|>\varepsilon_0\}}$ almost surely. Moreover, $T(\boldsymbol{\xi},\mathbf{x}_1) \leqslant \|\boldsymbol{\xi}\|$ almost surely, since $\mathrm{Lip}(f) \leqslant 1$. Let $a_0 := f(\mathbf{x}_0) - t^*$ and $a_1 := f(\mathbf{x}_1) - t^*$, and

$a_0, a_1 > 0$. Therefore,

$$
\begin{aligned}
I^p = & \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \Big\{ (1-\pi)\mathbb{E}\big[\big(a_0 + T(\boldsymbol{\xi}, \mathbf{x}_0)\big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_0 \big] + \pi \mathbb{E}\big[\big(a_1 + T(\boldsymbol{\xi}, \mathbf{x}_1)\big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_1 \big] \Big\} \\
\leqslant & \sup_{\mathbb{E}[\|\boldsymbol{\xi}\|^p] \leqslant \varepsilon^p} \Big\{ (1-\pi)\mathbb{E}\big[\big(a_0 + k_0 \|\boldsymbol{\xi}\| \mathbb{1}_{\{\|\boldsymbol{\xi}\| \leqslant \varepsilon_0\}} + \|\boldsymbol{\xi}\| \mathbb{1}_{\{\|\boldsymbol{\xi}\| > \varepsilon_0\}} \big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_0 \big] \\
& \qquad + \pi \mathbb{E}\big[\big(a_1 + \|\boldsymbol{\xi}\|\big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_1 \big] \Big\} \\
= & \sup_{\mathbb{E}[V^p] \leqslant \varepsilon^p} \Big\{ (1-\pi)\mathbb{E}\big[\big(a_0 + k_0 V \mathbb{1}_{\{V \leqslant \varepsilon_0\}} + V \mathbb{1}_{\{V > \varepsilon_0\}} \big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_0 \big] \\
& \qquad + \pi \mathbb{E}\big[\big(a_1 + V\big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_1 \big] \Big\} \\
= & \sup_{\mathbb{E}[V^p] \leqslant \varepsilon^p} \Big\{ (1-\pi) I_1 + \pi I_2 \Big\}.
\end{aligned}
\tag{A58}
$$

where the first equality uses the law of total expectation. Here, we write

$$
I_1 := \mathbb{E}\big[\big(a_0 + k_0 V \mathbb{1}_{\{V \leqslant \varepsilon_0\}} + V \mathbb{1}_{\{V > \varepsilon_0\}} \big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_0 \big] \text{ and } I_2 := \mathbb{E}\big[\big(a_1 + V\big)^p \mid \boldsymbol{\zeta} = \mathbf{x}_1 \big].
$$

Applying Lemma A8 conditionally, there exists $c_0 \in (0, \infty)$, depending only on (p, a_0, ε_0), such that

$$
I_1 \leqslant \Big(a_0 + k_0 \big(\mathbb{E}[V^p \mid \boldsymbol{\zeta} = \mathbf{x}_0] \big)^{1/p} + c_0 \mathbb{E}[V^p \mid \boldsymbol{\zeta} = \mathbf{x}_0] \Big)^p. \tag{A59}
$$

Moreover, by Lemma A6(i) (Jensen's inequality),

$$
I_2 \leqslant \Big(a_1 + \big(\mathbb{E}[V^p \mid \boldsymbol{\zeta} = \mathbf{x}_1] \big)^{1/p} \Big)^p. \tag{A60}
$$

Note that $\mathbb{E}[V^p] = (1-\pi)\mathbb{E}[V^p \mid \boldsymbol{\zeta} = \mathbf{x}_0] + \pi \mathbb{E}[V^p \mid \boldsymbol{\zeta} = \mathbf{x}_1]$. Let $v_0 := (\mathbb{E}[V^p \mid \boldsymbol{\zeta} = \mathbf{x}_0])^{1/p}$ and $v_1 := (\mathbb{E}[V^p \mid \boldsymbol{\zeta} = \mathbf{x}_1])^{1/p}$. Applying (A59) and (A60) to (A58) then yields

$$
\begin{aligned}
I^p \leqslant & \sup_{(1-\pi) v_0^p + \pi v_1^p \leqslant \varepsilon^p} \Big\{ (1-\pi)\big(a_0 + k_0 v_0 + c_0 v_0^p\big)^p + \pi \big(a_1 + v_1\big)^p \Big\} \\
= & \sup_{(1-\pi) u_0^p + \pi u_1^p \leqslant 1} \Big\{ (1-\pi)\big(a_0 + k_0 u_0 \varepsilon + c_0 u_0^p \varepsilon^p\big)^p + \pi \big(a_1 + u_1 \varepsilon\big)^p \Big\},
\end{aligned}
\tag{A61}
$$

where the equality follows from the change of variables $u_0 := v_0/\varepsilon$ and $u_1 := v_1/\varepsilon$. Applying Lemma A9 to (A61), there exist $L < 1$ and $M < \infty$ such that, for all $\varepsilon \in [0, \bar{\varepsilon}]$,

$$
I \leqslant \big((1-\pi) a_0^p + \pi a_1^p \big)^{1/p} + L\varepsilon + M(\varepsilon^2 + \varepsilon^p). \tag{A62}
$$

Let $\delta := (1-L)/2$. Then there exists $\eta > 0$ such that $M(\eta + \eta^{p-1}) < \delta$. Hence, for all sufficiently small $\varepsilon \leqslant \min\{\bar{\varepsilon}, \eta\}$, applying (A62) to (A57) yields

$$\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \mathcal{H}_p(f(\boldsymbol{\zeta} + \boldsymbol{\xi})) \leqslant t^* + c\big((1-\pi)a_0^p + \pi a_1^p\big)^{1/p} + cL\varepsilon + cM(\varepsilon + \varepsilon^{p-1})\varepsilon$$
$$< \mathcal{H}_p(f(\boldsymbol{\zeta})) + c(L+\delta)\varepsilon < \mathcal{H}_p(f(\boldsymbol{\zeta})) + c\varepsilon, \tag{A63}$$

where t^* satisfies (A56), namely, $\mathcal{H}_p(f(\boldsymbol{\zeta})) = t^* + c((1-\pi)a_0^p + \pi a_1^p)^{1/p}$. Consequently, (A63) contradicts (A52), which completes the proof. □

Proof of Theorem 6. For the "if" part, suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is of the form (11). Then the claim follows directly from Theorem 2 together with Theorem 5 of Wu et al. (2022).

For the "only if" part, we assume that there exists $c_f \geqslant 0$ such that (33) holds for any $F_0 \in \mathcal{M}(\mathbb{R}^n)$ and $\varepsilon > 0$. If $c_f = 0$, then (33) reduces to

$$\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \rho_h(f(\boldsymbol{\xi} + \boldsymbol{\zeta})) = \rho_h^{F_0}(f(\boldsymbol{\zeta})). \tag{A64}$$

An argument analogous to that for the case $c_f = 0$ in the proof of Theorem 3 shows that f is a constant function. Hence, f satisfies (11). Now consider the case $c_f > 0$. By the positive homogeneity of ρ_h, we assume without loss of generality that $c_f = 1$. In this case, (33) is equivalent to

$$\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \rho_h(f(\boldsymbol{\xi} + \boldsymbol{\zeta})) = \rho_h^{F_0}(f(\boldsymbol{\zeta})) + \varepsilon\|h'\|_q. \tag{A65}$$

We next show that f satisfies (11). To see this, we first show that f is Lipschitz continuous with $\mathrm{Lip}(f) \leqslant 1$. Notably, for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, set $F_0 = \delta_{\mathbf{x}}$ and $\varepsilon := \|\mathbf{y} - \mathbf{x}\|$. Then

$$f(\mathbf{y}) \leqslant \sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \rho_h(f(\boldsymbol{\xi} + \mathbf{x})) = f(\mathbf{x}) + \|h'\|_q \varepsilon = f(\mathbf{x}) + \|h'\|_q \|\mathbf{x} - \mathbf{y}\|.$$

Hence f is Lipschitz continuous with $\mathrm{Lip}(f) \leqslant \|h'\|_q < \infty$. We next show that $\mathrm{Lip}(f) \leqslant 1$. Note that when $p = \infty$, we have $\|h'\|_1 = 1$. Therefore, it suffices to consider the case $p \in (1, \infty)$. Suppose otherwise that $\mathrm{Lip}(f) > 1$. We claim that for any sufficiently small $\delta > 0$, there exist $\mathbf{x}_0 \in \mathbb{R}^n$, $t_0 > 0$, and $\boldsymbol{v}_0 \in \mathbb{R}^n$ with $\|\boldsymbol{v}_0\| = 1$ such that

$$f(\mathbf{x}_0 + t\boldsymbol{v}_0) - f(\mathbf{x}_0) \geqslant (1 + 2\delta)t, \quad \forall\, t \in (0, t_0]. \tag{A66}$$

Indeed, by the definition of $\mathrm{Lip}(f)$, for any sufficiently small $\delta > 0$ there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{y}$ such that $|f(\mathbf{x}) - f(\mathbf{y})|/\|\mathbf{x} - \mathbf{y}\| \geqslant 1 + 4\delta$. Without loss of generality, assume $f(\mathbf{y}) > f(\mathbf{x})$. Let $d := \|\mathbf{y} - \mathbf{x}\|$, define $\boldsymbol{v}_0 := (\mathbf{y} - \mathbf{x})/d$, and consider the function $\phi : [0, d] \to \mathbb{R}$ given by $\phi(s) := f(\mathbf{x} + s\boldsymbol{v}_0)$. Since f is Lipschitz, ϕ is absolutely continuous and hence differentiable almost everywhere, with $\phi(d) - \phi(0) = \int_0^d \phi'(s)\mathrm{d}s$. The inequality $f(\mathbf{y}) - f(\mathbf{x})/d \geqslant 1 + 4\delta$ implies that there exists $s_\delta \in (0, d)$ such that $\phi'(s_\delta) \geqslant 1 + 3\delta$. By the definition of the derivative, there exists $t_0 > 0$ such that $\phi(s_\delta + t) - \phi(s_\delta) \geqslant (1 + 2\delta)t$ for all $t \in (0, t_0]$. Setting $\mathbf{x}_0 := \mathbf{x} + s_\delta \boldsymbol{v}_0$, we obtain (A66), which proves the claim. Next, let U be a uniform random variable on $[0, 1]$. Fix $\delta > 0$ sufficiently small, and let $\mathbf{x}_0$, $\boldsymbol{v}_0$, and t_0 be such that (A66) holds. For any $\varepsilon > 0$ sufficiently small, define

$$R_\varepsilon := \varepsilon \frac{(h'(U))^{q/p}}{\|h'\|_q^{q/p}}, \quad \widetilde{R}_\varepsilon := \min\{R_\varepsilon, t_0\} \quad \text{and} \quad \boldsymbol{\xi}_1 := \boldsymbol{v}_0 \widetilde{R}_\varepsilon.$$

One can verify that $\mathbb{E}\|\boldsymbol{\xi}_1\|^p \leqslant \mathbb{E}[R_\varepsilon^p] = \varepsilon^p$. Note that $\lim_{\varepsilon \to 0} \mathbb{E}[(h'(U))^q \mathbb{1}_{\{R_\varepsilon > t_0\}}] = 0$ by the dominated convergence theorem. Hence, for any $\eta < 1 - 1/(1 + 2\delta)$, there exists $\varepsilon_\eta > 0$ such that

$$\mathbb{E}[(h'(U))^q \mathbb{1}_{\{R_\varepsilon > t_0\}}] \leqslant \eta \|h'\|_q^q, \qquad \forall\, \varepsilon \in (0, \varepsilon_\eta]. \tag{A67}$$

For $\varepsilon < \varepsilon_\eta$ sufficiently small, we then have

$$\begin{aligned}
&\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p \leqslant \varepsilon^p} \rho_h(f(\mathbf{x_0} + \boldsymbol{\xi})) \qquad\qquad \text{(A68)}\\
&\geqslant \rho_h\left(f(\mathbf{x}_0 + \boldsymbol{\xi}_1)\right)\\
&= \rho_h\left(f(\mathbf{x}_0 + \boldsymbol{v}_0 \widetilde{R}_\varepsilon)\right)\\
&\geqslant \rho_h\left(f(\mathbf{x}_0) + (1 + 2\delta)\widetilde{R}_\varepsilon\right)\\
&= f(\mathbf{x}_0) + (1 + 2\delta)\rho_h\left(\widetilde{R}_\varepsilon\right)\\
&= f(\mathbf{x}_0) + (1 + 2\delta)\mathbb{E}[\min\{R_\varepsilon, t_0\} h'(U)]\\
&\geqslant f(\mathbf{x}_0) + (1 + 2\delta)\left(\mathbb{E}[R_\varepsilon h'(U)] - \mathbb{E}[R_\varepsilon h'(U)\mathbb{1}_{\{R_\varepsilon > t_0\}}]\right)\\
&= f(\mathbf{x}_0) + (1 + 2\delta)\left(\varepsilon\|h'\|_q - \frac{\varepsilon}{\|h'\|_q^{q/p}}\mathbb{E}[(h'(U))^q \mathbb{1}_{\{R_\varepsilon > t_0\}}]\right)\\
&\geqslant f(\mathbf{x}_0) + (1 + 2\delta)(1 - \eta)\varepsilon\|h'\|_q > f(\mathbf{x}_0) + \varepsilon\|h'\|_q, \qquad\qquad \text{(A69)}
\end{aligned}$$

where the second inequality follows from (A66) and the fact that $\widetilde{R}_\varepsilon \leqslant t_0$ almost surely, the third equality uses the comonotonicity of $\widetilde{R}_\varepsilon$ and $h'(U)$, the fourth equality follows from the definition of R_ε, the fourth inequality follows from (A67), and the strict inequality holds because $(1+2\delta)(1-\eta) >$

1. Thus, (A69) contradicts (A65). It follows that f is Lipschitz continuous with $\mathrm{Lip}(f) \leqslant 1$. We next show that f satisfies (11). Since $\mathrm{Lip}(f) \leqslant 1$, suppose for contradiction that there exist $\tilde{\mathbf{x}} \in \mathbb{R}^n$ and $\varepsilon_0 > 0$ such that

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} \big(f(\tilde{\mathbf{x}} + \mathbf{y}) - f(\tilde{\mathbf{x}})\big) < \varepsilon_0. \tag{A70}$$

We next consider two cases: $p = \infty$ and $p \in (1, \infty)$. When $p = \infty$, we have

$$\sup_{\text{ess-sup}(\|\boldsymbol{\xi}\|) \leqslant \varepsilon_0} \rho_h\big(f(\tilde{\mathbf{x}} + \boldsymbol{\xi})\big) \leqslant \sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} f(\tilde{\mathbf{x}} + \mathbf{y}) < f(\tilde{\mathbf{x}}) + \varepsilon_0, \tag{A71}$$

where the first inequality follows from the monotonicity of ρ_h. Thus, (A71) contradicts (A65). Now suppose $p \in (1, \infty)$. Define

$$k := \sup_{\|\mathbf{y}\| \leqslant \varepsilon_0} \frac{f(\tilde{\mathbf{x}} + \mathbf{y}) - f(\tilde{\mathbf{x}})}{\varepsilon_0}.$$

By the strict inequality in (A70), we have $k < 1$. By Lemma A4, there exist $k_1 \in (k, 1)$, $\mathbf{x}_1 \in \mathbb{R}^n$, and $\varepsilon_1 > 0$ such that

$$f(\mathbf{x}_1 + \mathbf{y}) - f(\mathbf{x}_1) \leqslant k_1 \|\mathbf{y}\|, \quad \forall\, \|\mathbf{y}\| \leqslant \varepsilon_1. \tag{A72}$$

Next, note that $\lim_{t \to 0} \left(\int_{1-t}^1 (h'(u))^q \,\mathrm{d}u\right)^{1/q} = 0$ by the dominated convergence theorem. Hence, for any $\eta > 0$, there exists $t_\eta > 0$ such that

$$\left(\int_{1-t}^1 (h'(u))^q \mathrm{d}u\right)^{1/q} \leqslant \eta \|h'\|_q, \quad \forall\, t \in (0, t_\eta]. \tag{A73}$$

Define

$$\mathcal{X}_p = \left\{F^{-1}(U) : \int_0^1 |F^{-1}(u)|^p \mathrm{d}u < \infty,\ F^{-1}(U) \geqslant 0\right\}.$$

Then $\{F_X : X \in L^p\} = \{F_X : X \in \mathcal{X}_p\}$, and all random variables in $\mathcal{X}_p$ are comonotonic. Fix $\eta \in (0, 1 - k_1)$ and choose $t_\eta > 0$ such that (A73) holds. For $\varepsilon \in (0, \varepsilon_1 t_\eta^{1/p}]$ and any $X \in \mathcal{X}_p$ with $\|X\|_p \leqslant \varepsilon$, Markov's inequality yields

$$\mathbb{P}(X > \varepsilon_1) \leqslant \frac{\mathbb{E}[X^p]}{\varepsilon_1^p} \leqslant \left(\frac{\varepsilon}{\varepsilon_1}\right)^p \leqslant t_\eta. \tag{A74}$$

Consequently, for such η and ε, we have

$$
\begin{aligned}
&\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p}\rho_h(f(\mathbf{x}_1+\boldsymbol{\xi}))\\
&=\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p}\left\{\rho_h\left(f\left(\mathbf{x}_1+\boldsymbol{\xi}\right)\mathbb{1}_{\{\|\boldsymbol{\xi}\|\leqslant\varepsilon_1\}}+f\left(\mathbf{x}_1+\boldsymbol{\xi}\right)\mathbb{1}_{\{\|\boldsymbol{\xi}\|>\varepsilon_1\}}\right)\right\}\\
&\leqslant\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p}\left\{\rho_h\left(\left(f(\mathbf{x}_1)+k_1\|\boldsymbol{\xi}\|\right)\mathbb{1}_{\{\|\boldsymbol{\xi}\|\leqslant\varepsilon_1\}}+\left(f\left(\mathbf{x}_1\right)+\|\boldsymbol{\xi}\|\right)\mathbb{1}_{\{\|\boldsymbol{\xi}\|>\varepsilon_1\}}\right)\right\}\\
&=f\left(\mathbf{x}_1\right)+\sup_{\mathbb{E}\|\boldsymbol{\xi}\|^p\leqslant\varepsilon^p}\left\{\rho_h\left(k_1\|\boldsymbol{\xi}\|\mathbb{1}_{\{\|\boldsymbol{\xi}\|\leqslant\varepsilon_1\}}+\|\boldsymbol{\xi}\|\mathbb{1}_{\{\|\boldsymbol{\xi}\|>\varepsilon_1\}}\right)\right\}\\
&=f\left(\mathbf{x}_1\right)+\sup_{\mathbb{E}[X^p]\leqslant\varepsilon^p,\,X\geqslant0}\left\{\rho_h\left(k_1X\mathbb{1}_{\{X\leqslant\varepsilon_1\}}+X\mathbb{1}_{\{X>\varepsilon_1\}}\right)\right\}\\
&=f\left(\mathbf{x}_1\right)+\sup_{\mathbb{E}[X^p]\leqslant\varepsilon^p,\,X\in\mathcal{X}_p}\left\{\rho_h\left(k_1X\mathbb{1}_{\{X\leqslant\varepsilon_1\}}+X\mathbb{1}_{\{X>\varepsilon_1\}}\right)\right\}\\
&=f\left(\mathbf{x}_1\right)+\sup_{\mathbb{E}[X^p]\leqslant\varepsilon^p,\,X\in\mathcal{X}_p}\left\{\mathbb{E}[(k_1X\mathbb{1}_{\{X\leqslant\varepsilon_1\}}+X\mathbb{1}_{\{X>\varepsilon_1\}})h'(U)]\right\}\\
&\leqslant f\left(\mathbf{x}_1\right)+k_1\sup_{\mathbb{E}[X^p]\leqslant\varepsilon^p,\,X\in\mathcal{X}_p}(\mathbb{E}[(h'(U)\mathbb{1}_{\{X\leqslant\varepsilon_1\}})^q])^{1/q}(\mathbb{E}[X^p])^{1/p}\\
&\qquad+\sup_{\mathbb{E}[X^p]\leqslant\varepsilon^p,\,X\in\mathcal{X}_p}(\mathbb{E}[(h'(U)\mathbb{1}_{\{X>\varepsilon_1\}})^q])^{1/q}(\mathbb{E}[X^p])^{1/p}\\
&\leqslant f\left(\mathbf{x}_1\right)+k_1\varepsilon\sup_{\mathbb{E}[X^p]\leqslant\varepsilon^p,\,X\in\mathcal{X}_p}(\mathbb{E}[(h'(U)\mathbb{1}_{\{X\leqslant\varepsilon_1\}})^q])^{1/q}\\
&\qquad+\varepsilon\sup_{\mathbb{E}[X^p]\leqslant\varepsilon^p,\,X\in\mathcal{X}_p}(\mathbb{E}[(h'(U)\mathbb{1}_{\{X>\varepsilon_1\}})^q])^{1/q}\\
&\leqslant f\left(\mathbf{x}_1\right)+k_1\left\|h'\right\|_q\varepsilon+\left(\int_{1-t_\eta}^1(h'(u))^q\mathrm{d}u\right)^{1/q}\varepsilon\\
&\leqslant f\left(\mathbf{x}_1\right)+(k_1+\eta)\left\|h'\right\|_q\varepsilon<f\left(\mathbf{x}_1\right)+\left\|h'\right\|_q\varepsilon.
\end{aligned}
\tag{A75}
$$

Here, the first inequality follows from $\mathrm{Lip}(f)\leqslant1$ together with (A72). The fourth equality uses the law invariance of ρ_h. The fifth equality follows from the comonotonicity of $k_1X\mathbb{1}_{\{X\leqslant\varepsilon_1\}}+X\mathbb{1}_{\{X>\varepsilon_1\}}$ and $h'(U)$ for $X\in\mathcal{X}_p$. The second inequality follows from Hölder's inequality. The fourth inequality uses the nonnegativity of h', as h is a convex distortion function, together with (A74). The fifth inequality follows from (A73), and the final strict inequality holds because $k_1+\eta<1$. Thus, (A75) leads to a contradiction to (A65), thereby establishing (11). This completes the proof. □

Before proving Proposition 7, we need the following lemma.

Lemma A10. *For any $p\in[1,\infty), F_0\in\mathcal{M}_p\left(\mathbb{R}^n\right)$ and $\varepsilon\geqslant0$, let $\ell:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}_+$ satisfy $\ell(\mathbf{z},t)$ is* *convex in $t \in \mathbb{R}$ and there exists $M > 0$ such that*

$$|\ell(\mathbf{z}_1, t) - \ell(\mathbf{z}_2, t)| \leqslant M\|\mathbf{z}_1 - \mathbf{z}_2\|, \quad \forall, \mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^n,\, t \in \mathbb{R}.$$

We have the following statements hold.

(i) If $\lim_{t\to-\infty} \partial\ell(\mathbf{z}, t)/\partial t < -1$ for all $\mathbf{z} \in \mathbb{R}^n$, then we have

$$\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \inf_{t\in\mathbb{R}} \left\{t + \left(\mathbb{E}^F\left[\ell^p(\boldsymbol{\xi}, t)\right]\right)^{1/p}\right\} = \inf_{t\in\mathbb{R}} \sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \left\{t + \left(\mathbb{E}^F\left[\ell^p(\boldsymbol{\xi}, t)\right]\right)^{1/p}\right\}.$$

(ii) If $\lim_{t\to-\infty} \partial\ell(\mathbf{z}, t)/\partial t < 0 < \lim_{t\to\infty} \partial\ell(\mathbf{z}, t)/\partial t$ for all $z \in \mathbb{R}$, then we have

$$\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \inf_{t\in\mathbb{R}} \mathbb{E}^F\left[\ell^p(\boldsymbol{\xi}, t)\right] = \inf_{t\in\mathbb{R}} \sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \mathbb{E}^F\left[\ell^p(\boldsymbol{\xi}, t)\right].$$

Proof. (i) Denote by $\pi_{1,\ell}(F, t) := t + \left(\mathbb{E}^F\left[\ell^p(\boldsymbol{\xi}, t)\right]\right)^{1/p}$. With the similar arguments as in the proof of Lemma EC.8 in Wu et al. (2022), one can verify that $\pi_{1,\ell}(F, t)$ is concave in F for all $t \in \mathbb{R}$ and convex in t for all $F \in \mathcal{M}_p(\mathbb{R}^n)$. Moreover, we have $\lim_{t\to\pm\infty} \pi_{1,\ell}(F, t) = \infty$ for all $F \in \mathcal{M}_p(\mathbb{R}^n)$. Thus, the set of all minimizers of the problem $\inf_{t\in\mathbb{R}} \pi_{1,\ell}(F, t)$ is a closed interval. Denote by $t(F) := \inf \arg\min_t \pi_{1,\ell}(F, t)$. We will show that $\{t(F) : F \in \mathbb{B}_p(F_0, \varepsilon)\}$ is a subset of a compact set. For any $F \in \mathbb{B}_p(F_0, \varepsilon)$ and $t \in \mathbb{R}$, let $\boldsymbol{\xi} \sim F$ and $\boldsymbol{\zeta} \sim F_0$ such that $\mathbb{E}[\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p] \leqslant \varepsilon^p$, and we have

$$\begin{aligned}
|\pi_{1,\ell}(F, t) - \pi_{1,\ell}(F_0, t)| &= \left|\left(\mathbb{E}^F\left[\ell^p(\boldsymbol{\xi}, t)\right]\right)^{1/p} - \left(\mathbb{E}^{F_0}\left[\ell^p(\boldsymbol{\zeta}, t)\right]\right)^{1/p}\right| \\
&\leqslant \left(\mathbb{E}\left[|\ell(\boldsymbol{\xi}, t) - \ell(\boldsymbol{\zeta}, t)|^p\right]\right)^{1/p} \\
&\leqslant \left(\mathbb{E}\left[M^p\|\boldsymbol{\xi} - \boldsymbol{\zeta}\|^p\right]\right)^{1/p} \leqslant M\varepsilon,
\end{aligned} \tag{A76}$$

where the first inequality follows from the triangle inequality, and the last step uses the definition of the Wasserstein ball $\mathbb{B}_p(F_0, \varepsilon)$. Hence,

$$\pi_{1,\ell}(F, t(F_0)) \leqslant \pi_{1,\ell}(F_0, t(F_0)) + M\varepsilon. \tag{A77}$$

Note that $\pi_{1,\ell}(F_0, t) \to \infty$ as $t \to \pm\infty$. There exists $\Delta > 0$ such that $\pi_{1,\ell}(F_0, t) > \pi_{1,\ell}(F_0, t(F_0)) + 2M\varepsilon$ for all $t \notin [t(F_0) - \Delta, t(F_0) + \Delta]$. This, combined with (A76), imply that

$$\pi_{1,\ell}(F, t) \geqslant \pi_{1,\ell}(F_0, t) - M\varepsilon > \pi_{1,\ell}(F_0, t(F_0)) + M\varepsilon,\, \forall t \notin [t(F_0) - \Delta, t(F_0) + \Delta].$$

This, together with (A77), implies $\{t(F) : F \in \mathbb{B}_p(F_0, \varepsilon)\} \subseteq [t(F_0) - \Delta, t(G_0) + \Delta]$. Using a minimax theorem (see e.g., Sion (1958)), it holds that

$$\begin{aligned}\sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \inf_{t\in\mathbb{R}} \pi_{1,\ell}(F,t) &= \sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \inf_{t\in[t(F_0)-\Delta,t(F_0)+\Delta]} \pi_{1,\ell}(F,t) \\ &= \inf_{t\in[t(F_0)-\Delta,t(F_0)+\Delta]} \sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \pi_{1,\ell}(F,t) \\ &\geqslant \inf_{t\in\mathbb{R}} \sup_{F\in\mathbb{B}_p(F_0,\varepsilon)} \pi_{1,\ell}(F,t)\end{aligned}$$

The converse direction is trivial. Hence, we complete the proof.

(ii) The proof is similar to (i). □

Proof of Proposition 7. Without loss of generality, assume that $\mathrm{Lip}(f) = 1$. Denote by $\widetilde{\ell}_1(\mathbf{z}, t) := \ell_1(f(\mathbf{z}), t)$. For any $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^n$ with $\mathbf{z}_1 \neq \mathbf{z}_2$,

$$\begin{aligned}|\widetilde{\ell}_1(\mathbf{z}_1,t) - \widetilde{\ell}_1(\mathbf{z}_2,t)| &= |\ell_1(f(\mathbf{z}_1),t) - \ell_1(f(\mathbf{z}_2),t)| \\ &\leqslant b|f(\mathbf{z}_1) - f(\mathbf{z}_2)| \leqslant b\|\mathbf{z}_1 - \mathbf{z}_2\|, \quad \forall\, t \in \mathbb{R},\end{aligned}$$

where the first inequality follows from the uniform Lipschitz continuity of $\ell_1(z,t)$ in $z \in \mathbb{R}$, and the second inequality follows from the Lipschitz continuity of f. Thus, $\widetilde{\ell}_1(\mathbf{z},t)$ is Lipschitz continuous in $\mathbf{z}$ for all t, with constant b. Moreover, one can verify that $\widetilde{\ell}_1(\mathbf{z},t)$ is convex in t and satisfies $\lim_{t\to-\infty} \partial\widetilde{\ell}_1(\mathbf{z},t)/\partial t < 0 < \lim_{t\to\infty} \partial\widetilde{\ell}_1(\mathbf{z},t)/\partial t$ for all $\mathbf{z} \in \mathbb{R}^n$. Therefore, $\widetilde{\ell}_1(\mathbf{z},t)$ satisfies the assumptions of Lemma A10(ii). Hence,

$$\begin{aligned}\sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \mathcal{H}_1^F(f(\boldsymbol{\xi})) &= \sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \inf_{t\in\mathbb{R}} \mathbb{E}^F\left[\widetilde{\ell}_1(\boldsymbol{\xi},t)\right] \\ &= \inf_{t\in\mathbb{R}} \sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \mathbb{E}^F\left[\widetilde{\ell}_1(\boldsymbol{\xi},t)\right] \\ &= \inf_{t\in\mathbb{R}} \sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \mathbb{E}^F\left[\ell_1(f(\boldsymbol{\xi}),t)\right]. \end{aligned} \tag{A78}$$

We next show that

$$\sup_{F\in\mathbb{B}_1(F_0,\varepsilon)} \mathbb{E}^F\left[\ell_1(f(\boldsymbol{\xi}),t)\right] = \sup_{\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|]\leqslant\varepsilon} \mathbb{E}\left[\ell_1(f(\boldsymbol{\xi}),t)\right] = \mathbb{E}^{F_0}\left[\ell_1(f(\boldsymbol{\zeta}),t)\right] + b\varepsilon.$$

It suffices to establish the second equality. First,

$$\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|]\leqslant\varepsilon} \left(\mathbb{E}\left[\ell_1(f(\boldsymbol{\xi}),t)\right] - \mathbb{E}\left[\ell_1(f(\boldsymbol{\zeta}),t)\right]\right) \leqslant b \sup_{\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|]\leqslant\varepsilon} \mathbb{E}\|\boldsymbol{\xi}-\boldsymbol{\zeta}\| \leqslant b\varepsilon, \tag{A79}$$

where the first inequality follows from the Lipschitz continuity of $\ell_1(\cdot,t)$ and f. For the reverse inequality, we follow arguments similar to those used in Lemma A5 and Theorem 4. For the sake of completeness, we nevertheless provide a detailed proof. By (20), with similar arguments as in the proof of Lemma A5, there exist nondecreasing sequences $\{k_j\}_{j\in\mathbb{N}}$ and $\{m_j\}_{j\in\mathbb{N}}$ such that $m_j \to \infty$ as $j \to \infty$ and $(f(\mathbf{x}_0 + m_j \boldsymbol{v}_{k_j}) - f(\mathbf{x}_0))/m_j \geqslant 1 - \frac{1}{j}$. Let U be a uniform random variable on $[0,1]$, independent of $\boldsymbol{\zeta}$, and define

$$\boldsymbol{\xi}_j := \boldsymbol{\zeta} + m_j \boldsymbol{v}_{k_j} \mathbb{1}_{\{U\in A_j\}}, \quad \text{where} \quad A_j := [1-\varepsilon/m_j, 1).$$

One can verify that $\mathbb{E}\,\|\boldsymbol{\xi}_j - \boldsymbol{\zeta}\| = \varepsilon$. For each realization of $\boldsymbol{\zeta}$, denote by $\Delta_j := f(\boldsymbol{\zeta} + m_j \boldsymbol{v}_{k_j}) - f(\boldsymbol{\zeta})$. By Lipschitz continuity of f,

$$m_j(1-\frac{1}{j}) - 2\|\boldsymbol{\zeta} - \mathbf{x}_0\| \leqslant \Delta_j \leqslant m_j + 2\|\boldsymbol{\zeta} - \mathbf{x}_0\|. \tag{A80}$$

Thus $\Delta_j/m_j \to 1$ and $\Delta_j \to \infty$. Recall that for each $t \in \mathbb{R}$ there exists $z_0(t)$ such that $\lim_{m\to\infty} \frac{\ell_1(z_0(t)+m,t)-\ell_1(z_0(t),t)}{m} = b$, which implies that $\lim_{m\to\infty} \frac{\ell_1(z+m,t)-\ell_1(z,t)}{m} = b$ for all $z, t \in \mathbb{R}$. Note that $\Delta_j \to \infty$, as $j \to \infty$. Then, we have $\lim_{j\to\infty} \frac{\ell_1(f(\boldsymbol{\zeta})+\Delta_j,t)-\ell_1(f(\boldsymbol{\zeta}),t)}{\Delta_j} = b$. Therefore, for any $t \in \mathbb{R}$ and sufficiently large $m_j > \max\{1, \varepsilon\}$,

$$\begin{aligned}
&\sup_{\mathbb{E}[\|\boldsymbol{\xi}-\boldsymbol{\zeta}\|]\leqslant\varepsilon} \mathbb{E}\left[\ell_1(f(\boldsymbol{\xi}),t) - \ell_1(f(\boldsymbol{\zeta}),t)\right] \\
&\geqslant \mathbb{E}\left[\ell_1(f(\boldsymbol{\xi}_j),t) - \ell_1(f(\boldsymbol{\zeta}),t)\right] \\
&= \mathbb{E}\left[\left(\ell_1(f(\boldsymbol{\zeta}+m_j\boldsymbol{v}_{k_j}),t) - \ell_1(f(\boldsymbol{\zeta}),t)\right)\mathbb{1}_{A_j}\right] \\
&= \mathbb{E}\left[m_j \frac{\Delta_j}{m_j} \frac{(\ell_1(f(\boldsymbol{\zeta})+\Delta_j,t) - \ell_1(f(\boldsymbol{\zeta}),t))}{\Delta_j} \mathbb{1}_{A_j}\right] \\
&= \mathbb{E}\left[\frac{\Delta_j}{m_j} \frac{(\ell_1(f(\boldsymbol{\zeta})+\Delta_j,t) - \ell_1(f(\boldsymbol{\zeta}),t))}{\Delta_j}\right] \mathbb{E}\left[m_j \mathbb{1}_{A_j}\right] \\
&= \varepsilon\mathbb{E}\left[\frac{\Delta_j}{m_j} \frac{(\ell_1(f(\boldsymbol{\zeta})+\Delta_j,t) - \ell_1(f(\boldsymbol{\zeta}),t))}{\Delta_j}\right] \to b\varepsilon \text{ as } j\to\infty,
\end{aligned} \tag{A81}$$

where the third equality uses the independence of U and $\boldsymbol{\zeta}$, and the limit follows from the dominated convergence theorem, since

$$\left|\frac{\Delta_j}{m_j} \frac{(\ell_1(f(\boldsymbol{\zeta})+\Delta_j,t) - \ell_1(f(\boldsymbol{\zeta}),t))}{\Delta_j}\right| \leqslant b\frac{|\Delta_j|}{m_j} \leqslant b(1+2\|\boldsymbol{\zeta}-\mathbf{x}_0\|),$$

with the last inequality from (A80) and $m_j > 1$. Combining (A81) with (A79) and (A78), we obtain (36). This completes the proof. □

Proof of Proposition 8. Without loss of generality, assume $\mathrm{Lip}(f) = 1$. Note that $\rho_h \in \mathcal{R}_{\mathrm{coh}}$. Since f satisfies (20), Theorem 4(*i*) yields

$$\sup_{F \in \mathbb{B}_1(F_0,\varepsilon)} \rho_h^F\big(f(\boldsymbol{\xi})\big) = \sup_{G \in \mathcal{C}_1(f|F_0,\varepsilon)} \rho_h^G(X). \tag{A82}$$

Moreover, it follows from Proposition 2 in Pflug et al. (2012) (see also Theorem Wu et al. (2022)) that

$$\sup_{G \in \mathcal{C}_1(f|F_0,\varepsilon)} \rho_h^G(X) = \rho_h^{F_0}\big(f(\boldsymbol{\zeta})\big) + \varepsilon \|h'\|_\infty. \tag{A83}$$

Combining (A82) and (A83) completes the proof. □

Proof of Proposition 9. Note that we can rewrite $\mathrm{ex}_\alpha^F(X) = \max\{x \in \mathbb{R} : \mathbb{E}^F[\ell_\alpha(X - x)] \geqslant 0\}$, where $\ell_\alpha(x) := \alpha x_+ - (1-\alpha)x_-$, $x \in \mathbb{R}$. Denote by $c_f = \mathrm{Lip}(f)$. Therefore we have

$$\begin{aligned}
\sup_{G \in \mathcal{C}_1(f|F_0,c_f\varepsilon)} \mathrm{ex}_\alpha^G(X) &= \sup_{G \in \mathcal{C}_1(f|F_0,c_f\varepsilon)} \max\left\{x \in \mathbb{R} : \mathbb{E}^G[\ell_\alpha(X - x)] \geqslant 0\right\} \\
&= \max\{x \in \mathbb{R} : \sup_{G \in \mathcal{C}_1(f|F_0,c_f\varepsilon)} \mathbb{E}^G[\ell_\alpha(X - x)] \geqslant 0\} \\
&= \max\left\{x : \mathbb{E}^{F_0}[\ell_\alpha(f(\boldsymbol{\zeta}) - x)] + \alpha c_f \varepsilon \geqslant 0\right\},
\end{aligned}$$

where the last equality follows from the regularization result of a convex Lipschitz continuous function over a Wasserstein ball, i.e., $\sup_{G \in \mathbb{B}_1(G_0,\varepsilon)} \mathbb{E}^G[\ell_\alpha(X - x)] = \mathbb{E}^{G_0}[\ell_\alpha(X - x)] + \alpha\varepsilon$. This implies

$$\sup_{F \in \mathbb{B}_1(F_0,\varepsilon)} \mathrm{ex}_\alpha^F\left(f\left(\boldsymbol{\xi}\right)\right) = \max\left\{x : \mathbb{E}^{F_0}[\ell_\alpha(f(\boldsymbol{\zeta}) - x)] + \alpha c_f \varepsilon \geqslant 0\right\},$$

or equivalently, the unique solution to $\mathbb{E}^{F_0}[\ell_\alpha(f(\boldsymbol{\zeta}) - x)] + \alpha c_f \varepsilon = 0$. This completes the proof. □

A.3 Proofs for Section 5

Proof of Proposition 10. By Theorem 1, it is straightforward to verify the implications $(ii) \Rightarrow (i)$ and $(iii) \Rightarrow (ii)$. We only give the proof of $(i) \Rightarrow (iii)$.

For $(i) \Rightarrow (iii)$, suppose that there exists $c_f \geqslant 0$ such that (39) holds for any $\varepsilon > 0$ and $F_0 \in \mathcal{M}(\Xi)$. If $c_f = 0$, take $\rho = \mathrm{VaR}_\alpha$ with $\alpha \in [0,1)$. Let F_0 be such that $(Y_0, \mathbf{X}_0) \sim F_0$, where $Y_0 \equiv 1$ almost surely and $F_{\mathbf{X}_0}$ denote the marginal distribution of $\mathbf{X}_0$. Then, with similar arguments as in the proof of $(i) \Rightarrow (iii)$ in Theorem 1, it follows that f must be constant. Now consider $c_f > 0$ and take $\rho = \mathrm{VaR}_\alpha$ with $\alpha \in [0,1)$. Assume without loss of generality that $c_f = 1$. Let F_0 be such that $(Y_0, \mathbf{X}_0) \sim F_0$, where $Y_0 \equiv 1$ almost surely and $\mathbf{X}_0 \sim F_{\mathbf{X}_0}$. In this case, (39) reduces to

$$\sup_{F \in \mathbb{B}_p(F_{\mathbf{X}_0}, \varepsilon)} \mathrm{VaR}_\alpha^F(f(\mathbf{X})) = \sup_{G \in \mathcal{C}_p(f | F_{\mathbf{X}_0}, \varepsilon)} \mathrm{VaR}_\alpha^G(X),$$

which holds for any $\varepsilon > 0$ and $F_{\mathbf{X}_0} \in \mathcal{M}(\mathbb{R}^n)$. By Lemma A1 and the proof of Theorem 1 $(i) \Rightarrow (iii)$, it follows that

$$\sup_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) = \varepsilon, \tag{A84}$$

for any $\mathbf{x} \in \mathbb{R}^n$ and $\varepsilon > 0$. Similarly, let F_0 be such that $(Y_0, \mathbf{X}_0) \sim F_0$, where $Y_0 \equiv -1$ almost surely and $\mathbf{X}_0 \sim F_{\mathbf{X}_0}$. Then, for any $\varepsilon > 0$ and any $F_{\mathbf{X}_0} \in \mathcal{M}(\mathbb{R}^n)$, it holds that

$$\sup_{F \in \mathbb{B}_p(F_{\mathbf{X}_0}, \varepsilon)} \mathrm{VaR}_\alpha^F(-f(\mathbf{X})) = \sup_{G \in \mathcal{C}_p(-f | F_{\mathbf{X}_0}, \varepsilon)} \mathrm{VaR}_\alpha^G(X).$$

By Lemma A1, we have

$$f(\mathbf{x}) - \inf_{\|\mathbf{y}\| \leqslant \varepsilon} f(\mathbf{x} + \mathbf{y}) = \varepsilon, \quad \forall\, \mathbf{x} \in \mathbb{R}^n,\ \varepsilon > 0. \tag{A85}$$

Combining (A84) and (A85), we have (9) holds. This completes the proof. □

Proof of Proposition 11. The "if" part is a direct consequence of $(iii) \Rightarrow (i)$ in Proposition 41 and the "only if" part has been already proved in the proof for $(i) \Rightarrow (iii)$ in Proposition 41. □

Proof of Corollary 2. By Propositions 1, 2, 10, and Theorem 6, the conclusion follows immediately. □

References

Aolaritei, L., Lanzetti, N., Chen, H., and Dörfler, F. (2023). Distributional uncertainty propagation via optimal transport *arXiv preprint arXiv:2205.00343*

Bellini, F., Klar, B., Müller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54, 41–8.

Blanchet, J., Kang, Y., and Murthy, K. (2019). Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, **56**(3), 830–857.

Chen, Z., Kuhn, D., and Wiesemann, W. (2024). Data-driven chance constrained programs over Wasserstein balls. *Operations Research*, **72**(1), 410–424.

Chen, Z., and Xie, W. (2021). Sharing the value-at-risk under distributional ambiguity. *Mathematical Finance*, **31**(1), 531–559.

Clarkson, J. A. (1936). Uniformly convex spaces. *Transactions of the American Mathematical Society*, **40**(3), 396-414.

Kuhn, D., Shafiee, S., and Wiesemann, W. (2025). Distributionally robust optimization. *Acta Numerica*, **34**, 579-804.

Föllmer, H., and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time*. Fourth Edition. Walter de Gruyter, Berlin.

Filipović, D. and Svindland, G. (2007). Convex risk measures on Lp. *Preprint, www. math. lmu. de/filipo/PAPERS/crmlp.*

Gao, R., Chen, X., and Kleywegt, A. J. (2024). Wasserstein distributionally robust optimization and variation regularization. *Operations Research*, **72**(3), 1177–1191.

Ho-Nguyen, N., and Wright, S. J. (2023). Adversarial classification via distributional robustness with Wasserstein ambiguity. *Mathematical Programming*, **198**(2), 1411-1447.

Huber, P. J. (1992). Robust estimation of a location parameter. *Breakthroughs in statistics: Methodology and distribution* (pp. 492-518). New York, NY: Springer New York

Koenker, R., and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.

Kuhn, D., Esfahani, P. M., Nguyen, V. A., and Shafieezadeh-Abadeh, S. (2019). Wasserstein distributionally robust optimization: Theory and applications in machine learning. *Operations Research and Management Science in the age of analytics*. Informs, 2019. 130-166.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in mathematical economics*. Tokyo: Springer Japan, 2001. 83-95.

Liu, F., Mao, T., Wang, R., and Wei, L. (2022). Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Mathematics of Operations Research*, **47**(3), 2494–2519.

Mao, T., Wang, R., and Wu, Q. (2022). Model aggregation for risk evaluation and robust optimization. *Management Science*, forthcoming.

Olea, J. L. M., Rush, C., Velez, A. and Wiesel, J. (2022). The out-of-sample prediction error of the square-root-LASSO and related estimators. *arXiv:2211.07608.*

Pflug, G. C., Pichler, A., and Wozabal, D. (2012). The 1/N investment strategy is optimal under high model ambiguity. *Journal of Banking and Finance*, **36**(2), 410–417.

Rieder, U. (1978). Measurable selection theorems for optimization problems. *manuscripta mathematica*, **24**(1), 115–131.

Rockafellar, R. T. *Convex Analysis.* (1970) Princeton University Press, Princeton, New Jersey.

Santambrogio, F. (2015). Optimal transport for applied mathematicians.

Shafieezadeh-Abadeh, S., Kuhn, D., Esfahani, P.M. (2019) Regularization via mass transportation. *Journal of Machine Learning Research*, **20**(103), 1–68

Shapiro, A. (2013). On Kusuoka representation of law invariant risk measures. *Mathematics of Operations*

Research, **38**(1), 124–152.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Sion, M. (1958). On general minimax theorems. *Pacific Journal of Mathematics*, **8**(1), 171–176.

Wu, Q., Li, J. Y. M., and Mao, T. (2022). On generalization and regularization via wasserstein distributionally robust optimization. *Management Science*, forthcoming.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, 819–847.

Xie, W. (2021). On distributionally robust chance constrained programs with Wasserstein distance. *Mathematical Programming*, **186**(1), 115–155.

Yaari, M.E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.