Risk Quadrangle and Robust Optimization Based on φ-Divergence

Cheng Peng* Anton Malandii* Stan Uryasev*

Abstract

This paper studies robust and distributionally robust optimization based on the extended φ-divergence under the Fundamental Risk Quadrangle framework. We present the primal and dual representations of the quadrangle elements: risk, deviation, regret, error, and statistic. The framework provides an interpretation of portfolio optimization, classification and regression as robust optimization. We furnish illustrative examples demonstrating that many common problems are included in this framework. The φ-divergence risk measure used in distributionally robust optimization is a special case. We conduct a case study to visualize the risk envelope.

Keywords: Risk Quadrangle, φ-Divergence, φ-Divergence Risk Measure, Entropic Value-at-Risk, Conditional Value-at-Risk, Distributionally Robust Optimization.

*Department of Applied Mathematics and Statistics, State University of New York, Stony Brook, NY 11794, USA. Correspondence: cheng.peng.1@stonybrook.edu

1 Introduction

1.1 Demonstrating Examples

We start the introduction with two demonstrating examples.

Example A The first is an interpretation of Markowitz portfolio optimization [Markowitz, 1952], Large Margin Distribution Machine [Zhang and Zhou, 2014], and least squares regression as robust loss minimization. Robust optimization in this study refers to minimizing the maximum weighted loss, where the weight comes from an uncertainty set. For the first example, we define the uncertainty set $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$ of random variables Q as a Euclidean ball of radius $\sqrt{\beta}$ centered at 1 with expected value of 1 :

$$\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta} = \{Q \in \mathcal{L}^2 : \mathbb{E}[Q] = 1, \mathbb{E}[\varphi(Q)] \leq \beta\}, \quad \varphi(x) = (x-1)^2. \tag{1.1}$$

Set $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$ will appear in portfolio optimization, classification, and regression problems. Denote by $\sigma(X)$ the standard deviation of a random variable X. Consider a random portfolio loss $X(\boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{l}$, where $\boldsymbol{w} \in \mathbb{R}^d$ is a vector of portfolio weights and $\boldsymbol{l}$ is a random vector of negative asset returns. Let $\mathbf{1} = (1, \dots, 1)^\top \in \mathbb{R}^d$. As an example of the general equivalence in Section 8, the following two problems

have the same optimal objective function value and the set of solution vectors:

Markowitz portfolio optimization	**Robust expected loss minimization**
$$\min_{\mathbf{1}^\top \boldsymbol{w}=1} \mathbb{E}[X(\boldsymbol{w})] + \sqrt{\beta}\sigma(X(\boldsymbol{w})), \tag{1.2}$$	$$\min_{\mathbf{1}^\top \boldsymbol{w}=1} \max_{Q \in \mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[QX(\boldsymbol{w})] \,. \tag{1.3}$$

Problem (1.3) is the robust version of the expected loss minimization problem $\min_{\mathbf{1}^\top \boldsymbol{w}=1} \mathbb{E}[X(\boldsymbol{w})]$.

The following is an interpretation of the Large Margin Distribution Machine classification algorithm as a robust optimization. Consider an attribute (random vector of features) $\boldsymbol{X}$, label Y, and decision vector $\boldsymbol{w}$. The margin is defined by $L(\boldsymbol{w}, b) = Y(\boldsymbol{w}^\top \boldsymbol{X} - b)$. Denote by $\gamma(\boldsymbol{w})$ a regularization term. The following two problems have the same optimal objective function value and the set of solution vectors:

Large Margin Distribution Machine	**Robust expected margin maximization**
$$\min_{\boldsymbol{w},b} \mathbb{E}[-L(\boldsymbol{w}, b)] + \sqrt{\beta}\sigma(-L(\boldsymbol{w}, b)) + \gamma(\boldsymbol{w}), \tag{1.4}$$	$$\min_{\boldsymbol{w},b} \max_{Q \in \mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[-QL(\boldsymbol{w}, b)] + \gamma(\boldsymbol{w}) \,. \tag{1.5}$$

Problem (1.5) is the regularized robust version of the expected margin maximization problem $\max_{\boldsymbol{w}} \mathbb{E}[L(\boldsymbol{w}, b)]$.

The following is an interpretation of least squares regression as robust optimization. Consider a dependent variable (regressant) Y, a vector of regressors (factors) $\boldsymbol{X} = (X_1, \dots, X_d)$, a class of functions $\mathcal{F}$, and an intercept $C \in \mathbb{R}$. The regression residual is defined by $Z_f = Y - f(\boldsymbol{X}) - C$, and the residual without the intercept C is defined by $\bar{Z}_f = Y - f(\boldsymbol{X})$. Denote by $\|X\|_2$ the Euclidean norm of a random variable X. The following two problems have the same optimal solution (f, C):

Least squares regression

$$\min_{f\in\mathcal{F},C\in\mathbb{R}} ||Z_f||_2 \ , \tag{1.6}$$

Deviation minimization

$$\min_{f\in\mathcal{F}} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[Q\bar{Z}_f] - \mathbb{E}[\bar{Z}_f] \tag{1.7}$$

$$\text{calculate } C = \mathbb{E}[\bar{Z}_f] \ . \tag{1.8}$$

The considered approach is based on the Fundamental Risk Quadrangle (FRQ) framework [Rockafellar and Uryasev, 2013] connecting different measures associated with risk and uncertainty. The basic elements forming a risk quadrangle are four functionals of a random variable X, error $\mathcal{E}(X)$, regret $\mathcal{V}(X)$, deviation $\mathcal{D}(X)$, and risk $\mathcal{R}(X)$. Statistic $\mathcal{S}(X)$ binds these four functionals. The quadrangle elements satisfy the following relations

$$\mathcal{V}(X) = \mathcal{E}(X) + \mathbb{E}[X] \ , \quad \mathcal{R}(X) = \mathcal{D}(X) + \mathbb{E}[X] \ , \tag{1.9}$$

$$\underset{C\in\mathbb{R}}{\operatorname{argmin}}\{C + \mathcal{V}(X - C)\} = \mathcal{S}(X) = \underset{C\in\mathbb{R}}{\operatorname{argmin}}\{\mathcal{E}(X - C)\} \ , \tag{1.10}$$

$$\mathcal{R}(X) = \min_{C\in\mathbb{R}}\{C + \mathcal{V}(X - C)\}, \quad \mathcal{D}(X) = \min_{C\in\mathbb{R}} \mathcal{E}(X - C) \ . \tag{1.11}$$

A regression problem minimizes an error $\mathcal{E}(Z_f)$ of the residual Z_f and estimates the conditional statistic $\mathcal{S}(Y|\boldsymbol{X})$. A portfolio optimization problem minimizes risk $\mathcal{R}(X)$ or deviation $\mathcal{D}(X)$ of the portfolio loss X (minimization of risk is equivalent to the minimization of deviation with the expectation constraint). Regret $\mathcal{V}(X)$ is an anti-utility function frequently used in stochastic optimization.

The following mean quadrangle (Example 1, Rockafellar and Uryasev [2013]) relates Markowitz portfolio optimization, Large Margin Distribution Machine and least squares regression.

Mean Quadrangle

$$\mathcal{R}(X) = \mathbb{E}[X] + \lambda\sigma(X) = \text{ safety margin tail risk,}$$
$$\mathcal{V}(X) = \mathbb{E}[X] + \lambda\left\|X\right\|_2 = \mathcal{L}^2\text{-regret, scaled,}$$
$$\mathcal{D}(X) = \lambda\sigma(X) = \text{ standard deviation, scaled,}$$
$$\mathcal{E}(X) = \lambda\left\|X\right\|_2 = \mathcal{L}^2\text{-error, scaled,}$$
$$\mathcal{S}(X) = \mathbb{E}[X] = \text{ mean.}$$

The interpretation of (1.3) and (1.7) as a robust optimization is obtained from the dual representation of risk and deviation in the following mean quadrangle

$$\mathcal{R}(X) = \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[QX],$$
$$\mathcal{V}(X) = \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta}} \mathbb{E}[QX],$$
$$\mathcal{D}(X) = \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[QX] - \mathbb{E}[X],$$
$$\mathcal{E}(X) = \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta}} \mathbb{E}[QX] - \mathbb{E}[X],$$
$$\mathcal{S}(X) = \mathbb{E}[X],$$

where $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$ is defined in (1.1) and the uncertainty set $\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta}$ of random variables Q removes the condition $\mathbb{E}[Q]=1$ in (1.1)

$$\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta} = \{Q \in \mathcal{L}^2 : \mathbb{E}[\varphi(Q)] \leq \beta\}, \quad \varphi(x) = (x-1)^2. \tag{1.12}$$

Example B The next example shows the relation between CVaR optimization [Rockafellar and Uryasev, 2000], ν-support vector machine [Schölkopf et al., 2000], quantile regression [Koenker and Bassett Jr, 1978] and robust optimization. Let $\nu = 1-\alpha$. The equivalence of ν-SVM and CVaR optimization was studied by Gotoh and Takeda [2004]; Takeda and Sugiyama [2008]. Define the uncertainty set $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$

$$\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta} = \{Q \in \mathcal{L}^2 \mid \mathbb{E}[Q]=1, \mathbb{E}[\varphi(Q)] \leq \beta\}, \quad \varphi(x) = \begin{cases} 0, & x \in [0, \frac{1}{1-\alpha}] \\ +\infty, & \text{otherwise} \end{cases}. \tag{1.13}$$

$\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$ will appear in portfolio optimization, classification, and regression problems. Similarly to Example A, in each of the following three pairs of problems, the optimizations on the left and right have the same optimal objective function value and the set of solution vectors:

CVaR portfolio optimization

$$\min_{\mathbf{1}^T\boldsymbol{w}=1} \mathrm{CVaR}_\alpha(X(\boldsymbol{w})) , \tag{1.14}$$

Robust loss minimization

$$\min_{\mathbf{1}^T\boldsymbol{w}=1} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[QX(\boldsymbol{w})] , \tag{1.15}$$

ν-SVM

$$\min_{\boldsymbol{w},b} \mathrm{CVaR}_\alpha(-L(\boldsymbol{w},b)) + \gamma(\boldsymbol{w}), \tag{1.16}$$

Robust expected margin maximization

$$\min_{\boldsymbol{w}} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[-QL(\boldsymbol{w},b)] + \gamma(\boldsymbol{w}) . \tag{1.17}$$

Quantile regression

$$\min_{f\in\mathcal{F}, C\in\mathbb{R}} \mathcal{E}_\alpha(Z_f) , \tag{1.18}$$

Deviation minimization

$$\min_{f\in\mathcal{F}} \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[Q\bar{Z}_f] - \mathbb{E}[\bar{Z}_f] \tag{1.19}$$

$$\text{calculate } C \in \mathrm{VaR}_\alpha[\bar{Z}_f] , \tag{1.20}$$

where $X_+ = \max\{0, X\}$, $X_- = \max\{0, -X\}$, and $\mathcal{E}(X) = \left[\frac{\alpha}{1-\alpha}X_+ + X_-\right]$ is the normalized Koenker-Bassett error.

CVaR portfolio optimization, ν-SVM, and quantile regression are connected by the quantile quadrangle (Example 2, Rockafellar and Uryasev [2013]). The interpretation as robust optimization is obtained from the dual representation, which is presented below together with the primal representation.

Quantile Quadrangle

$$\begin{aligned}
\mathcal{R}_\alpha(X) &= \mathrm{CVaR}_\beta(X) = \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[QX], \\
\mathcal{V}_\alpha(X) &= \frac{1}{1-\beta}\mathbb{E}[X_+] = \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta}} \mathbb{E}[QX], \\
\mathcal{D}_\alpha(X) &= \mathrm{CVaR}_\alpha(X) - \mathbb{E}[X] = \max_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[QX] - \mathbb{E}[X], \\
\mathcal{E}_\alpha(X) &= \mathbb{E}\left[\frac{\alpha}{1-\alpha}X_+ + X_-\right] = \max_{Q\in\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta}} \mathbb{E}[QX], \\
\mathcal{S}_\alpha(X) &= \mathrm{VaR}_\alpha(X) ,
\end{aligned}$$

where the uncertainty set $\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta}$ is defined by

$$\mathcal{Q}^{\mathcal{V}}_{\varphi,\beta} = \{Q \in \mathcal{L}^2 : \mathbb{E}[\varphi(Q)] \leq \beta\}, \quad \varphi(x) = \begin{cases} 0, & x \in [0, \frac{1}{1-\alpha}] \\ +\infty, & \text{otherwise} \end{cases}. \tag{1.21}$$

The robust representations (1.15), (1.17), (1.19) are implied by the dual representations of risk and deviation in the quantile quadrangle.

Distributionally robust optimization minimizes the maximum expected loss, where the expectation is taken under any probability measure from an uncertainty set. Ben-Tal et al. [2013] studies the uncertainty set such that the φ-divergence [Csiszár, 1963; Morimoto, 1963] between the alternative measure P and the reference measure P_0, denoted by $D_\varphi(P||P_0)$, is not larger than a chosen value. We show that the dual formulations (1.15), (1.17), (1.19) in Example B can be interpreted as distributionally robust optimization, while the dual formulations (1.3), (1.5), (1.7) in Example A cannot be interpreted as such. The interpretation is based on the equivalent representation of risk and deviation measure in quantile quadrangle as distributionally robust expectation and distributionally robust deviation from expectation, respectively,

$$\mathcal{R}_\alpha(X) = \max_{P \in \mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[X] \tag{1.22}$$

$$\mathcal{D}_\alpha(X) = \max_{P \in \mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[X] - \mathbb{E}X, \tag{1.23}$$

where $\mathcal{P}_{\varphi,\beta}$ is an uncertainty set of probability measures with $\varphi(x)$ defined in (1.21)

$$\mathcal{P}_{\varphi,\beta} = \{P \in \mathcal{P}(\Sigma) : D_\varphi(P||P_0) \leq \beta\}.$$

The demonstrated relations are examples of a general framework, under which many other connections are revealed. For a general class of functions $\varphi(x)$, called the extended divergence function (see Definition 2.8), we derive the dual and primal representation of the risk quadrangle. Gotoh and Uryasev [2017] studies classification as risk minimization problem. The primal representation facilitates optimization problem statements. The dual representation of quadrangle elements provides an interpretation of portfolio optimization, classification and regression as robust optimization. The dual representation of regular risk measure is studied in Rockafellar and Uryasev [2013]. The dualization of coherent regret is studied in Sun et al. [2020]; Rockafellar [2020]; Fröhlich and Williamson [2022a,b]; Rockafellar [2023]. We apply the theorem in Sun et al. [2020] to prove certainty equivalence relation between dual regret measure and dual risk measure.

Furthermore, for a subclass of functions called divergence function (see Definition 2.9), we provide an interpretation of portfolio optimization, classification and regression as distributionally robust optimization. The risk measure in relevant risk quadrangle is proposed in Ahmadi-Javid [2011]. The dual representation of the risk measure is a special case of the representation of coherent risk measure in Proposition 4.1 in Artzner et al. [1999]. Shapiro [2017] studies distributionally robust optimization considering the law-invariant risk measures.

1.2 Paper Contributions

We propose an extended φ-divergence quadrangle and prove regularity of the quadrangle in dual representation (Section 3). We derive the primal representation of the quadrangle elements:

risk, regret, deviation, error, and statistic (Section 4). The dual representation provides an interpretation of portfolio optimization, classification, and regression problems as a robust optimization (Section 8). Compared with existing literature, this is a first study on robust interpretation of classification and regression. The envelopes for these three problems are identical, even if the envelopes are not defined by an extended φ-divergence, which can be of independent interest.

A specific case of the extended φ-divergence quadrangle called φ-divergence quadrangle (Section 3) provides an interpretation of portfolio optimization, classification, and regression problem as distributionally robust optimization (Section 8). The extension of such quadrangles builds a connection of well-known risk and error measures with distributionally robust optimization. For example, the χ^2-divergence risk measure is the second-order superquantile risk measure (Example 7), and the extended χ^2-divergence risk measure is safety margin tail risk used in Markowitz portfolio optimization (Example 2).

Additionally, we recover the underlying φ-divergence from quadrangle elements (Section 6). We study optimality conditions in primal and dual optimization problem (Section 5). We provide various examples for demonstrating the approach (Section 7). We visualize risk identifiers for Markowitz portfolio optimisation, Large Margin Distribution Machine and least squares regression (Section 9). The code for the case study is available for downloading.

2 Mathematical Preliminaries

This section presents key definitions and notations.

2.1 Functional Space Setting

This section discusses the functional space setting we adopt for further analysis.

Let (Ω, Σ, P_0) be a probability space, where P_0 is a reference measure, and let $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$ denote an extended set of real numbers. As can be seen from Section 1, we work with stochastic convex functionals, elements of the FRQ, which need to be defined on the appropriate space of random variables, real-valued measurable functions $X : \Omega \to \mathbb{R}$.

The choice of $\mathcal{L}^p := \mathcal{L}^p(\Omega, \Sigma, P_0),\ p \in [1, \infty)$ seems to be reasonable, however, one still has to be careful since if $\mathcal{R} : \mathcal{L}^p \to \overline{\mathbb{R}}$ is a proper convex risk measure, then either $\mathcal{R}(\cdot)$ is finite valued and continuous on $\mathcal{L}^p$ or $\mathcal{R}(X) = +\infty$ on a dense set of points $X \in \mathcal{L}^p$ (cf. [Shapiro et al., 2014, Proposition 6.8]). Therefore, for some risk measures, it may be even impossible to find an appropriate space. Moreover, it is not possible to construct a finite valued convex risk measure on a space larger than $\mathcal{L}^1$ (cf. [Shapiro et al., 2014, Proposition 6.31]).

Our study concentrates on so-called φ-divergence risk measures for which the natural choice of a functional space can be an Orlicz space paired with a divergence function satisfying

$$\varphi(0) < +\infty, \quad \lim_{x \to +\infty} \frac{\varphi(x)}{x} = +\infty, \tag{2.1}$$

suggested by Dommel and Pichler [2020] and adopted by Fröhlich and Williamson [2022b].

However, this particular choice of a space excludes important divergence functions such as the total variation distance (TVD), fitting in the framework of Shapiro [2017], which uses $\mathcal{L}^p$ in general and switches to $\mathcal{L}^\infty$ for certain divergence functions.

Of course, the simplest way would be to avoid the complications arising in the infinite-dimensional setting by working with finite Ω. Then every function $X : \Omega \to \mathbb{R}$ is measurable, and the space of all such functions can be identified with the Euclidean space. Such an approach was taken by Bayraksan and Love [2015].

In light of everything mentioned above, we take a safe path by following [Rockafellar and Uryasev, 2013] in taking $\mathcal{L}^2$ as our working space assuming finiteness where needed. This choice will also allow us to rely on the extensive theory developed for the FRQ in this setting.

2.2 The Fundamental Risk Quadrangle Framework

Let $X \in \mathcal{L}^2$ be a real-valued random variable. Mathematical expectation and variance of a random variable X with respect to the reference measure is denoted by $\mathbb{E}[X]$, and by $\mathbb{V}[X]$. The set of all probability measures on (Ω, Σ) is denoted by $\mathcal{P}(\Sigma)$.

A functional $\rho : \mathcal{L}^2 \to \overline{\mathbb{R}}$ is called *convex* if

$$\rho\left(\lambda X + (1-\lambda)Y\right) \leq \lambda\rho(X) + (1-\lambda)\rho(Y),\ \forall\, X, Y \in \mathcal{L}^2,\ \lambda \in [0,1],$$

and *closed* if

$$\left\{X \in \mathcal{L}^2 | \rho(X) \leq c\right\} \text{ is a closed set } \forall\, c < \infty.$$

Next, we introduce the essential elements in the Fundamental Risk Quadrangle Theory. The framework is proposed in Rockafellar and Uryasev [2013]. Rockafellar and Royset [2015] relaxes some technical conditions, which we follow in this study. For reader's convenience, we put more details on quadrangle theory in Appendix A.

Definition 2.1 (Regular Risk Measure)**.** A closed convex functional $\mathcal{R} : \mathcal{L}^2 \to \overline{\mathbb{R}}$ is called a *regular measure of risk* if it satisfies:

$$\mathcal{R}(C) = C,\ \forall\, C = const \qquad \mathcal{R}(X) > \mathbb{E}[X],\ \forall\, X \neq const$$

Definition 2.2 (Regular Deviation Measure)**.** A closed convex functional $\mathcal{D} : \mathcal{L}^2 \to \overline{\mathbb{R}}^+$ is called a *regular measure of deviation* if it satisfies:

$$\mathcal{D}(C) = 0,\ \forall\, C = const \qquad \mathcal{D}(X) > 0,\ \forall\, X \neq const$$

Definition 2.3 (Regular Regret Measure)**.** A closed convex functional $\mathcal{V} : \mathcal{L}^2 \to \overline{\mathbb{R}}$ is called a *regular measure of regret* if it satisfies the following axioms

$$\mathcal{V}(0) = 0, \qquad \mathcal{V}(X) > \mathbb{E}[X],\ \forall\, X \neq const$$

Definition 2.4 (Regular Error Measure)**.** A closed convex functional $\mathcal{E} : \mathcal{L}^2 \to \overline{\mathbb{R}}^+$ is called a *regular measure of error* if it satisfies the following axioms:

$$\mathcal{E}(0) = 0, \qquad \mathcal{E}(X) > 0,\ \forall\, X \neq const$$

Definition 2.5 (Risk Quadrangle)**.** A quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ of measures of risk, deviation, regret, and error satisfying the following relationships is called a *risk quadrangle*:

(Q1) **error projection:** $\mathcal{D}(X) = \inf_C \left\{ \mathcal{E}(X - C) \right\}$;

(Q2) **certainty equivalence:** $\mathcal{R}(X) = \inf_C \{ C + \mathcal{V}(X - C) \}$;

(Q3) **centerness:** $\mathcal{R}(X) = \mathcal{D}(X) + \mathbb{E}[X], \quad \mathcal{V}(X) = \mathcal{E}(X) + \mathbb{E}[X]$.

Moreover, the quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ is bound by the statistic $\mathcal{S}(X)$ satisfying:

$$\mathcal{S}(X) = \underset{C \in \mathbb{R}}{\operatorname{argmin}} \{ \mathcal{E}(X - C) \} = \underset{C \in \mathbb{R}}{\operatorname{argmin}} \{ C + \mathcal{V}(X - C) \}.$$

Definition 2.6 (Regular Risk Quadrangle)**.** A regular risk quadrangle is a risk quadrangle with regular elements $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$.

Remark 2.1 (Error Projection and Certainty Equivalence)**.** In order to establish the validity of a given quartet $(\mathcal{R}, \mathcal{D}, \mathcal{V}, \mathcal{E})$ as a quadrangle, it is sufficient to demonstrate the satisfaction of either conditions (Q1) and (Q3), or conditions (Q2) and (Q3), as conditions (Q1) and (Q2) are intrinsically linked through the condition (Q3). Indeed,

$$\mathcal{R}(X) = \inf_C \{ C + \mathcal{V}(X - C) \} = \inf_C \{ \mathcal{E}(X - C) \} + \mathbb{E}[X] = \mathcal{D}(X) + \mathbb{E}[X].$$

Let $Z_f = Y - f(X) - C$, $\bar{Z}_f = Y - f(X)$, where $C \in \mathbb{R}$, f belongs to a class of functions $\mathcal{F}$.

Definition 2.7 (Regression)**.** A regression problem is defined as

$$\min_{f \in \mathcal{F}, C} \mathcal{E}(Z_f) \ . \tag{2.2}$$

Theorem 2.1 (Error Shaping Decomposition of Regression (Theorem 3.2, Rockafellar et al. [2008]))**.** *The solution to regression in Definition 2.7 is characterized by the prescription that*

$$f, C \in \underset{f, C}{\operatorname{argmin}} \, \mathcal{E}(Z_f) \ \textit{ if and only if } \begin{cases} f \in \operatorname{argmin}_f \mathcal{D}(\bar{Z}_f) \\ C \in \mathcal{S}(\bar{Z}_f) \end{cases} . \tag{2.3}$$

2.3 Divergence and Related Risk Measure

Definition 2.8 (Extended Divergence Function)**.** A convex lower semi-continuous function $\varphi : \mathbb{R} \to \overline{\mathbb{R}}$ is an *extended divergence function* if

$$\varphi(1) = 0, \quad \operatorname{dom}(\varphi) = \mathbb{R}, \quad 1 \in \operatorname{int}(\{x : \varphi(x) < +\infty\}) \ , \tag{2.4}$$

where the interior is denoted by int.

Definition 2.9 (Divergence Function)**.** A *divergence function* $\varphi(x)$ is an extended divergence function in Definition 2.8 that additionally satisfies

$$\varphi(x) = +\infty \ \text{ for } \ x < 0 \ . \tag{2.5}$$

Definition 2.10 (φ-Divergence)**.** Consider probability measures P and P_0, where P is dominated by P_0. For a divergence function $\varphi(x)$, the *φ-divergence* of P from P_0 is defined by

$$D_\varphi(P||P_0) := \int_\Omega \varphi\left(\frac{dP}{dP_0}\right) dP_0 \ . \tag{2.6}$$

The integral can be equivalently written as $\mathbb{E}_{P_0}\left[\varphi\left(\frac{dP}{dP_0}\right)\right]$. The following φ-divergences are discussed in subsequent examples. The indicator divergence is defined by function, $\varphi(x) = \mathbf{1}_{[0,(1-\alpha)^{-1}]}(x)$, where $\varphi(x) = 0$ if $x \in [0,(1-\alpha)^{-1}]$ and $+\infty$ otherwise; for Pearson χ^2-divergence, $\varphi(x) = \frac{1}{2}(x-1)^2$; for total variation distance (which is also a divergence), $\varphi(x) = \frac{1}{2}|x-1|$; for Kullback-Leibler divergence, $\varphi(x) = x\log x$.

Definition 2.11 (Conjugate Functional, Risk Envelope, Risk Identifier [Rockafellar and Uryasev, 2013])**.** Let $\rho : \mathcal{L}^2 \to \overline{\mathbb{R}}$ be a closed convex functional. Then a functional $\rho^* : \mathcal{L}^2 \to \overline{\mathbb{R}}$ is said to be *conjugate* to ρ if

$$\rho(X) = \sup_{Q\in\mathcal{Q}}\{\mathbb{E}[XQ] - \rho^*(Q)\}, \quad \forall\, X \in \mathcal{L}^2, \tag{2.7}$$

where $\mathcal{Q} = \mathrm{dom}(\rho^*)$ is called the *risk envelope* associated with ρ, and Q furnishing the maximum in (2.7) is called a *risk identifier* for X.

Definition 2.12 (Generalized Entropic Risk Measure [Ben-Tal and Teboulle, 1987, 2007; Föllmer and Schied, 2011])**.** Consider a divergence function $\varphi(x)$ (Definition 2.9). The generalized entropic risk measure is defined by

$$\mathcal{R}_\varphi(X) = \sup_{P\in\mathcal{P}(\Sigma)}\{\mathbb{E}_P[X] - D_\varphi(P||P_0)\}. \tag{2.8}$$

The envelope representation is as follows

$$\mathcal{R}_\varphi(X) = \sup_{Q\in\mathcal{Q}_\varphi^{1,+}}\{\mathbb{E}[XQ] - \mathbb{E}[\varphi(Q)]\}, \tag{2.9}$$

$$\mathcal{Q}_\varphi^{1,+} = \{Q \in \mathcal{L}^2 : Q \ge 0, \mathbb{E}[Q] = 1\}. \tag{2.10}$$

By convex conjugacy of (2.9) (Theorem 4.4 of Ben-Tal and Teboulle [2007]),

$$\mathbb{E}[\varphi(Q)] = \sup_{X\in\mathcal{L}^2}\{\mathbb{E}[XQ] - \mathcal{R}_\varphi(X)\}\,. \tag{2.11}$$

The primal representation is as follows

$$\mathcal{R}_\varphi(X) = \inf_C\{C + \mathbb{E}[\varphi^*(X-C)]\}. \tag{2.12}$$

Definition 2.13 (φ-divergence risk measure [Dommel and Pichler, 2020])**.** Consider a divergence function $\varphi(x)$. The φ-divergence risk measure is defined by

$$\mathcal{R}_{\varphi,\beta}(X) = \sup_{P\in\mathcal{P}_{\varphi,\beta}}\mathbb{E}_P[X]\,, \tag{2.13}$$

$$\mathcal{P}_{\varphi,\beta} = \{P \in \mathcal{P}(\Sigma) : D_\varphi(P||P_0) \le \beta\}\,. \tag{2.14}$$

3 Dual Representation of φ-Divergence Quadrangle

This section introduces the dual representation of φ-divergence quadrangle, and proves the regularity of the quadrangle and the coherency of a special case of this quadrangle. As the section title implies, there is a primal φ-divergence quadrangle that will be introduced subsequently in Section 4.

Definition 3.1 (Dual Representation of Extended φ-Divergence Quadrangle)**.** For an extended φ-divergence function and $X \in \mathcal{L}^2$, the dual extended φ-divergence quadrangle is defined by

$$\mathcal{R}_{\varphi,\beta}(X) = \sup_{Q \in \mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[XQ] \,, \tag{3.1}$$

$$\mathcal{V}_{\varphi,\beta}(X) = \sup_{Q \in \mathcal{Q}_{\varphi,\beta}} \mathbb{E}[XQ] \,, \tag{3.2}$$

$$\mathcal{D}_{\varphi,\beta}(X) = \sup_{Q \in \mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[X(Q-1)] \,, \tag{3.3}$$

$$\mathcal{E}_{\varphi,\beta}(X) = \sup_{Q \in \mathcal{Q}_{\varphi,\beta}} \mathbb{E}[X(Q-1)] \,, \tag{3.4}$$

$$\mathcal{S}_{\varphi,\beta}(X) = \operatorname*{argmin}_{C \in \mathbb{R}} \sup_{Q \in \mathcal{Q}_{\varphi,\beta}} \mathbb{E}[(X-C)(Q-1)] \,, \tag{3.5}$$

where

$$\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta} = \{Q \in \mathcal{L}^2 : \mathbb{E}[Q] = 1, \mathbb{E}[\varphi(Q)] \le \beta\}, \tag{3.6}$$

$$\mathcal{Q}_{\varphi,\beta} = \{Q \in \mathcal{L}^2 : \mathbb{E}[\varphi(Q)] \le \beta\} \tag{3.7}$$

are the envelopes associated with $\mathcal{R}(X)$ and $\mathcal{V}(X)$ respectively.

Note that there is no requirement in the envelope that $Q \ge 0$.

Definition 3.2 (Dual φ-Divergence Quadrangle)**.** The dual extended φ-divergence quadrangle is called a dual φ-divergence quadrangle if the extended φ-divergence is φ-divergence.

We prove the dual representation of the extended φ-divergence quadrangle. The proof of aversity of the risk measure constructs a feasible random variable inspired by Ang et al. [2018]. The proof of the relation between risk and regret follows Sun et al. [2020]. Ang et al. [2018]; Sun et al. [2020] work with coherent risk measures. The proving techniques are of broader interest. Ang et al. [2018] proves that 1 being a relative interior point of the envelope $\mathcal{Q}$ is sufficient for a coherent risk measure to be risk averse. Sun et al. [2020] proves that removing $\mathbb{E}Q = 1$ in the envelope of coherent risk measure generates a coherent regret measure.

Theorem 3.1 (Dual Representation of Extended φ-Divergence Quadrangle)**.** *Let $\varphi(x)$ be an extended φ-divergence function, $X \in \mathcal{L}^2$. The quartet $(\mathcal{R}_{\varphi,\beta}, \mathcal{D}_{\varphi,\beta}, \mathcal{V}_{\varphi,\beta}, \mathcal{E}_{\varphi,\beta})$ defined by (3.1)–(3.4) is a regular risk quadrangle with the statistic (3.5).*

Proof. First, we verify the conditions for regular risk measure in Definition 2.1.

Closedness and Convexity: Since the envelope $\mathcal{Q}$ is closed and convex ([Rockafellar et al., 2006; Rockafellar and Uryasev, 2013]), then $\mathcal{R}_{\varphi,\beta}(X)$ is closed (lower semicontinuous) and convex as a maximum of continuous affine functions.

Constancy: Constancy is implied by the condition $\mathbb{E}Q = 1$,

$$\sup_{Q \in \mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \mathbb{E}[CQ] = \sup_{Q \in \mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} C\,\mathbb{E}[Q] = C \,.$$

Risk aversity: We can construct a Q_0 such that the strict inequality holds for $\mathcal{R}_{\varphi,\beta}(X) > \mathbb{E}[X]$. As a function of r, $P(X \le r)$ is a nondecreasing, right-continuous function with a range in $[0,1]$. Thus for a nonconstant X, there exists $r \in \mathbb{R}$, $p \in (0,1)$ such that $P(X \le r) = p$, $P(X > r) = 1 - p$. By convexity of $\varphi(x)$ and $1 \subset \operatorname{int}(\{x : \varphi(x) < +\infty\})$, there exists $\delta > 0$

such that $\varphi(x) \leq \beta$ for $x \in (1-\delta, 1+\delta)$. Then, there exists $\delta_1 \in (0,\delta)$, $\delta_2 \in (0,\delta)$ such that $\delta_1 = \frac{1-p}{p}\delta_2$. Define Q_0 by

$$Q_0(\omega) = \begin{cases} 1-\delta_1, & \omega : X(\omega) \leq r \\ 1+\delta_2, & \omega : X(\omega) > r \end{cases} . \tag{3.8}$$

The feasibility can be checked by $\mathbb{E}[\varphi(Q_0)] \leq \beta$, $\mathbb{E}Q_0 = 1$.

We have

$$\begin{aligned}
\mathbb{E}[XQ_0] =& \mathbb{E}[XQ_0|X \leq r]P(X \leq r) + \mathbb{E}[XQ_0|X > r] && (3.9)\\
=& p(1-\delta_1)\mathbb{E}[X|X \leq r] + (1-p)(1+\delta_2)\mathbb{E}[X|X > r] && (3.10)\\
=& p\mathbb{E}[X|X \leq r] + (1-p)\mathbb{E}[X|X > r] - p\delta_1\mathbb{E}[X|X \leq r] + (1-p)\delta_2\mathbb{E}[X|X > r] && (3.11)\\
=& \mathbb{E}[X] + p\delta_1(\mathbb{E}[X|X > r] - \mathbb{E}[X|X \leq r]) && (3.12)\\
>& \mathbb{E}[X] \, . && (3.13)
\end{aligned}$$

Thus $\mathcal{R}_{\varphi,\beta}(X)$ is a regular risk measure.

Next, we verify the conditions for regular regret measure.

Closedness and Convexity: Same with the proof above for regular risk measure.

Risk aversity: For $X \neq const$,

$$\mathcal{V}_{\beta,\varphi}(X) \geq \mathcal{R}_{\beta,\varphi}(X) > \mathbb{E}X. \tag{3.14}$$

The first inequality is due to $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta} \subset \mathcal{Q}_{\varphi,\beta}$.

Zeroness:

$$\sup_{Q \in \mathcal{Q}_{\varphi,\beta}} \mathbb{E}[0 \cdot Q] = 0. \tag{3.15}$$

The proof of Theorem 1 in Sun et al. [2020] (which works on coherent risk measure) can be applied here to show that a regular regret measure can be obtained by removing condition $\mathbb{E}Q = 1$ in (3.6). Thus the risk (3.1) and regret (3.2) satisfies (Q2) in Definition 2.5.

Deviation (3.3) and error (3.4) measure are obtained by centerness formulae (Q3) (see Definition 2.5). With Theorem A.1, we can show the regularity of deviation and error, and that the minimum in C for a regular regret measure is attainable. The optimal C is $\mathcal{S}_{\varphi,\beta}(X)$. □

Proposition 4.1 of Artzner et al. [1999] proves the coherency of risk measures that have representation $\sup_{P \in \mathcal{P}} \mathbb{E}_P[X]$ for any set $\mathcal{P}$. The setting in Artzner et al. [1999] is finite $\mathcal{R}(X)$ and finite Ω.

An alternative proof of risk (3.1) and regret (3.2) satisfying (Q2) in Definition 2.5 can be obtained from the primal representations in Section 4. The relation (Q2) can be directly observed from the primal risk and regret.

Consider the φ-divergence quadrangle in Definition 3.2. Since $\varphi(x) = +\infty$ for $x < 0$, $\mathbb{E}[\varphi(Q)] \leq \beta$ implies that $Q \geq 0$ almost surely. The envelope with $Q \geq 0$ is the necessary and sufficient condition for monotonicity of the convex homogeneous functional associated with such envelope [Rockafellar et al., 2006; Rockafellar and Uryasev, 2013]. $\mathcal{R}_{\varphi,\beta}(X)$ becomes the φ-divergence risk measure (Definition 2.13), which is a coherent risk measure. In fact, it is straightforward to verify that the quadrangle elements of φ-divergence quadrangle satisfy the axioms of coherent risk, regret, deviation and error measure defined in Fröhlich and Williamson [2022a].

We discuss the interpretation of Q in $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$. Consider the extended φ-divergence quadrangle. The covariance between random variables X and Q is $\operatorname{cov}(X,Q) = \mathbb{E}[(X - \mathbb{E}X)(Q - \mathbb{E}Q)]$. Since $\mathbb{E}Q = 1$ by (3.6), $\operatorname{cov}(X,Q) = \mathbb{E}[X(Q-1)]$. The deviation (3.3) can be written as $\sup_{Q\in\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}} \operatorname{cov}(X,Q)$. Thus the optimal Q^* tracks X as closely as possible. Next, we consider only φ-divergence quadrangle. Define indicator function $\mathcal{I}_A(x) = 1$ if $x \in A$ and 0 otherwise. For every $Q \in \mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$, we can define a probability measure on (Ω, Σ) by

$$P_Q(A) = \mathbb{E}[\mathcal{I}_A(\omega)Q(\omega)], \quad A \in \Sigma\,. \tag{3.16}$$

Q is the Radon–Nikodym derivative dP_Q/dP_0. Then the condition $\mathbb{E}[\varphi(Q)] \le \beta$ can be equivalently expressed by $D_\varphi(P||P_0) \le \beta$. The envelope $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$ has a one-to-one correspondence to a set of probability measures

$$\mathcal{P}_{\varphi,\beta} = \{P \in \mathcal{P}(\Sigma) : D_\varphi(P||P_0) \le \beta\}. \tag{3.17}$$

The dual representations (3.1) and (3.3) can be equivalently written as

$$\mathcal{R}_{\varphi,\beta}(X) = \sup_{P\in\mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[X] \tag{3.18}$$

$$\mathcal{D}_{\varphi,\beta}(X) = \sup_{P\in\mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[X] - \mathbb{E}[X]\,. \tag{3.19}$$

4 Primal Extended φ-Divergence Quadrangle

This section discusses the primal representation of the elements in the dual φ-divergence quadrangle.

Definition 4.1 (Primal Extended φ-Divergence Quadrangle)**.** For an extended divergence function $\varphi(x)$ and $X \in \mathcal{L}^2$, the Primal Extended φ-Divergence quadrangle is defined by

$$\mathcal{R}_{\varphi,\beta}(X) = \inf_{\substack{C\in\mathbb{R},\\ t>0}} t\left\{C + \beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - C\right)\right]\right\}, \tag{4.1}$$

$$\mathcal{D}_{\varphi,\beta}(X) = \inf_{\substack{C\in\mathbb{R},\\ t>0}} t\left\{C + \beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - C\right) - \frac{X}{t}\right]\right\}, \tag{4.2}$$

$$\mathcal{V}_{\varphi,\beta}(X) = \inf_{t>0} t\left\{\beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t}\right)\right]\right\}, \tag{4.3}$$

$$\mathcal{E}_{\varphi,\beta}(X) = \inf_{t>0} t\left\{\beta + \mathbb{E}\left[\varphi^*\left(\frac{X}{t}\right) - \frac{X}{t}\right]\right\}, \tag{4.4}$$

$$\mathcal{S}_{\varphi,\beta}(X) = \operatorname*{argmin}_{C\in\mathbb{R}} \inf_{t>0} t\left\{\frac{C}{t} + \beta + \mathbb{E}\left[\varphi^*\left(\frac{X-C}{t}\right)\right]\right\}. \tag{4.5}$$

Definition 4.2 (Primal φ-Divergence Quadrangle)**.** The primal extended φ-divergence quadrangle is called a primal φ-divergence quadrangle if the extended φ-divergence is φ-divergence.

Theorem 4.1 (Primal Extended φ-Divergence Quadrangle)**.** *Let $\varphi(x)$ be an extended divergence function, $X \in \mathcal{L}^2$. Elements of the dual extended φ-divergence quadrangle in Theorem 3.1 can be presented as (4.1)–(4.5) in Definition 4.1. The optimal t and C in (4.1)–(4.5) are attainable.*

Proof. Consider the regret (3.2)

$$\mathcal{V}_{\varphi,\beta} = \sup_{Q \in \mathcal{Q}_{\varphi,\beta}} \mathbb{E}[XQ] = -\inf_{Q \in \mathcal{Q}_{\varphi,\beta}} \mathbb{E}[-QX]. \tag{4.6}$$

Consider the Lagrangian dual problem of $\inf_{Q:Q\in\mathcal{Q}_{\varphi,\beta}} \mathbb{E}[-QX]$

$$\sup_{t\geq 0} \inf_{Q} \left\{\mathbb{E}[-XQ] + t\left(\mathbb{E}[\varphi(Q)] - \beta\right)\right\}. \tag{4.7}$$

Denote the optimal t by t^*. If $t^* = 0$, then

$$\sup_{t\geq 0} \inf_{Q} \left\{\mathbb{E}[-XQ] + t\left(\mathbb{E}[\varphi(Q)] - \beta\right)\right\} = \inf_{Q} \mathbb{E}[-XQ] = -\infty. \tag{4.8}$$

Thus for all $t \geq 0$,

$$\inf_{Q} \left\{\mathbb{E}[-XQ] + t\left(\mathbb{E}[\varphi(Q)] - \beta\right)\right\} = -\infty\ . \tag{4.9}$$

Thus if $t^* = 0$, the optimum is also attained at $t > 0$. If $t^* > 0$, $t > 0$ and $t \geq 0$ are the same for the problem. Thus, we can substitute $t \geq 0$ with $t > 0$ in the Lagrange dual problem.

Then,

$$\sup_{t>0} \inf_{Q} \left\{\mathbb{E}[-XQ] + t\left(\mathbb{E}[\varphi(Q)] - \beta\right)\right\} \tag{4.10}$$

$$= \sup_{t>0} \inf_{Q} (-t) \left\{\mathbb{E}\left[\frac{X}{t}Q - \varphi(Q)\right] + \beta\right\} \tag{4.11}$$

$$= -\inf_{t>0} \sup_{Q} t \left\{\mathbb{E}\left[\frac{X}{t}Q - \varphi(Q)\right] + \beta\right\}\ . \tag{4.12}$$

Next, we prove that

$$-\inf_{t>0} \sup_{Q} t \left\{\mathbb{E}\left[\frac{X}{t}Q - \varphi(Q)\right] + \beta\right\} = -\inf_{t>0} t \left\{\beta + \mathbb{E}\varphi^*\left(\frac{X}{t}\right)\right\}. \tag{4.13}$$

We consider two cases where the following condition is satisfied and not satisfied

$$\sup_{Q} \left\{\mathbb{E}\left[\frac{X}{t}Q - \varphi(Q)\right]\right\} < +\infty \quad \text{for some } t\ . \tag{4.14}$$

When (4.14) is satisfied, since $XQ/t - \varphi(Q)$ is a normal convex integrand [Shapiro, 2017], *sup* and expectation in (4.12) are exchangeable by Theorem 3A of Rockafellar [1976]. Thus, (4.13) holds.

When (4.14) is not satisfied, $\sup_Q\{\mathbb{E}[XQ/t - \varphi(Q)]\} = +\infty$ for all t. We have

$$-\inf_{t>0} \sup_{Q} t \left\{\mathbb{E}\left[\left(\frac{X}{t}\right) Q - \varphi(Q)\right] + \beta\right\} = -\infty.$$

We also have that

$$t\left(\mathbb{E}\varphi^*\left(\frac{X}{t}\right)+\beta\right) = t\left(\mathbb{E}\left[\sup_Q\left\{\left(\frac{X}{t}\right)Q-\varphi(Q)\right\}\right]+\beta\right) \tag{4.15}$$

$$\geq \sup_Q t\left\{\mathbb{E}\left[\left(\frac{X}{t}\right)Q-\varphi(Q)\right]+\beta\right\} \tag{4.16}$$

$$=+\infty. \tag{4.17}$$

Thus

$$-\inf_{t>0} t\left(\mathbb{E}\varphi^*\left(\frac{X}{t}\right)+\beta\right)=-\infty. \tag{4.18}$$

We see that (4.13) holds with or without the condition (4.14). With (4.10)–(4.12), (4.13), we obtain

$$\sup_{t>0}\inf_Q\left\{\mathbb{E}[-XQ]+t\left(\mathbb{E}[\varphi(Q)]-\beta\right)\right\}=-\inf_{t>0} t\left\{\beta+\mathbb{E}\varphi^*\left(\frac{X}{t}\right)\right\}. \tag{4.19}$$

Strong duality for the convex problem holds since the following Slater's condition is valid for $Q=1$

$$\exists Q\ :\ Q\in\mathcal{Q}_{\varphi,\beta},\ \mathbb{E}[\varphi(Q)]<\beta\ . \tag{4.20}$$

Thus

$$\mathcal{V}_{\varphi,\beta}=-\inf_{Q\in\mathcal{Q}_{\varphi,\beta}}\mathbb{E}[-QX]=\inf_{t>0} t\left\{\beta+\mathbb{E}\varphi^*\left(\frac{X}{t}\right)\right\}. \tag{4.21}$$

By regularity, the statistic $S_{\varphi,\beta}(X)$ is attainable. Denote the optimal C and t by C^* and t^*. If $t^*>0$, $\frac{S_{\varphi,\beta}(X)}{t^*}$ is attainable. We showed that if $t^*=0$, any $t>0$ is also optimal. $\frac{S_{\varphi,\beta}(X)}{t^*}$ is attainable. By change of variable, C^* in (4.1),(4.2) equals $\frac{S_{\varphi,\beta}(X)}{t^*}$. Thus t^* and C^* in (4.1)–(4.5) are attainable.

The primal representation of the other elements can be obtained similarly by Lagrange dual problem, or by direct calculation using the quadrangle relations in Definition 2.5. □

The primal representation of the risk measure (4.1) is studied in the literature under different technical conditions.

Fröhlich and Williamson [2022b] starts with the primal representation of coherent regret and obtains the coherent risk with (Q3) centerness relation in Definition 2.5.

5 Statistic and Risk Identifier

5.1 Statistic and Characterizing Equations

This section characterizes statistic function and studies optimality conditions for (C,t) in the primal problem in Theorem 4.1.

Lemma 5.1 (Convexity)**.** *Let $f : \mathbb{R} \times (0, \infty) \to \mathbb{R}$ be such that*

$$f(C,t) = C + t\beta + \mathbb{E}\Big[t\varphi^*\Big(\frac{X-C}{t}\Big)\Big]. \tag{5.1}$$

Then $f(C,t)$ is convex in (C,t) and

$$\partial_{(C,t)}(f(C,t)) = (1,\beta)^\top + \mathbb{E}\left[\partial_{(C,t)}\left(t\varphi^*\left(\frac{X-C}{t}\right)\right)\right], \tag{5.2}$$

where $\partial_{(C,t)}(f(C,t))$ denotes a subdifferential of a convex function $f(C,t)$ with respect to the vector $(C,t)^\top \in \mathbb{R} \times (0,\infty)$, cf. [Rockafellar, 1970, Definition 23.1]. The "+" sign in (5.2) is understood in the sense of the Minkowski sum.

Proof. To prove the first part of the lemma it suffices to establish that the function

$$\psi(z,t) = t\varphi^*(z/t), \quad z \in \mathbb{R},\ t \in (0,\infty)$$

is convex. This follows from the fact that the function $h(z,t) = tg(z/t),\ z \in \mathbb{R}^n,\ t > 0$ is convex if and only if g is convex. Such function h is called a *perspective function,* cf. [Dacorogna and Maréchal, 2008, Lemma 2.1]. Hence, since φ^* is convex then ψ is also convex as a perspective function. Therefore, $f(C,t)$ is convex since convexity is preserved under linear transformations.

The second part of the lemma follows from [Rockafellar, 1977, Theorem 23]. Indeed, since the function under the expectation in (5.1) is convex, hence measurable (cf. Rockafellar and Wets [1998]), the subdifferential can be interchanged with the expectation. $\square$

Theorem 5.1 (Characterization of $\mathcal{S}_{\varphi,\beta}$)**.** *Let $(\mathcal{R}_{\varphi,\beta}, \mathcal{D}_{\varphi,\beta}, \mathcal{V}_{\varphi,\beta}, \mathcal{E}_{\varphi,\beta})$ be a Primal Extended φ-Divergence Quadrangle. Statistic in this quadrangle equals*

$$\mathcal{S}_{\varphi,\beta}(X) = \left\{C \in \mathbb{R} : 0 \in (1,\beta)^\top + \mathbb{E}\left[\partial_{(C,t)}\left(t\varphi^*\left(\frac{X-C}{t}\right)\right)\right]\right\}. \tag{5.3}$$

Proof. Definition 2.5 implies that the statistic is equal to

$$\begin{aligned} \mathcal{S}_{\varphi,\beta}(X) &= \operatorname*{argmin}_{C\in\mathbb{R}}\{\, C + \mathcal{V}_{\varphi,\beta}(X-C)\,\} \\ &= \operatorname*{argmin}_{C\in\mathbb{R}} \inf_{t>0} f(C,t)\ , \end{aligned} \tag{5.4}$$

where $f(C,t) = C + t\beta + \mathbb{E}\Big[t\varphi^*\Big(\frac{X-C}{t}\Big)\Big]$. To find the statistic one has to minimize $f(C,t)$ with respect to (C,t). Since $f(C,t)$ is convex, cf. Lemma 5.1, then it reaches the minimum if and only if

$$0 \in \partial_{(C,t)} f(C,t)\ . \tag{5.5}$$

Therefore, cf. Lemma 5.1, condition (5.5) is equivalent to

$$0 \in (1,\beta)^\top + \mathbb{E}\left[\partial_{(C,t)}\left(t\varphi^*\left(\frac{X-C}{t}\right)\right)\right]. \tag{5.6}$$

$\square$

If for an extended divergence function $\varphi(x)$, the conjugate $\varphi^*(z)$ is positive homogeneous, then the expression (4.5) for statistic is reduced to

$$\mathcal{S}_{\varphi,\beta}(X) = \underset{C\in\mathbb{R}}{\operatorname{argmin}}\ C + \mathbb{E}[\varphi^*(X-C)]\ . \tag{5.7}$$

The [Rockafellar and Uryasev, 2013, Expectation Theorem] in this case implies

$$\mathcal{S}_{\varphi,\beta}(X) = \left\{ C \in \mathbb{R} : \mathbb{E}\left[\frac{\partial^-}{\partial z}\varphi^*(z)\bigg|_{x=X-C} \right] \le 1 \le \mathbb{E}\left[\frac{\partial^+}{\partial z}\varphi^*(z)\bigg|_{x=X-C} \right] \right\}, \tag{5.8}$$

where $\dfrac{\partial^-}{\partial z}, \dfrac{\partial^+}{\partial z}$ denote left and right derivatives with respect to $z \in \mathbb{R}$. As a finite convex homogeneous function, $\varphi^*(z)$ is the support function of a closed interval (Corollary 13.2.2, Rockafellar [1970]). The convex conjugate of a support function is an indicator function. Since $\varphi(1) = 0$, it must be in the form of the $\varphi(x)$ in Example 8.

In fact, Dommel and Pichler [2020] provided optimality conditions for (C, t) in (4.1). For differentiable function φ^*, they developed a set of equations known as the *characterizing equations* for optimal (C, t). Further, we provide a system of equations similar to the characterizing equations developed by Dommel and Pichler [2020].

Definition 5.1 (Characterizing Equations)**.** Let $\varphi^*(z) \in C^1(\mathbb{R})$. Characterizing system of equations is defined by:

$$\begin{cases} \mathbb{E}\left[\dfrac{d\varphi^*(z)}{dz}\bigg|_{z=\frac{X-C}{t}} \right] = 1\ , \\[2ex] \beta + \mathbb{E}\left[\varphi^*\left(\dfrac{X-C}{t} \right) \right] - \dfrac{1}{t}\mathbb{E}\left[(X-C)\dfrac{d\varphi^*(z)}{dz}\bigg|_{z=\frac{X-C}{t}} \right] = 0\ . \end{cases} \tag{5.9}$$

The following Corollary 5.1 provides an expression for the statistic $\mathcal{S}_{\varphi,\beta}$ with smooth $\varphi^*(z)$.

Corollary 5.1 (Characterization of $\mathcal{S}_{\varphi,\beta}$: Smooth Case)**.** *Let $\varphi^*(z) \in C^1(\mathbb{R})$, then the statistic equals*

$$\mathcal{S}_{\varphi,\beta}(X) = \{C \in \mathbb{R} : (C, t) \textit{ is a solution to Characterizing Equations } (5.9)\ \}.$$

Proof. Replacing the subdifferential in (5.6) with the gradient $\nabla_{(C,t)}$ leads to the system of equations (5.9). □

5.2 Optimality Conditions for Risk Identifier

Lemma 5.2 (Subgradients of expectation, Bauschke and Combettes [2011])**.** *Let $(\Omega, \mathcal{A}, P_0)$ be a probability space and $\psi : \mathbb{R} \to \overline{\mathbb{R}}$ be a proper, lsc, and convex function. Set*

$$\rho = \mathbb{E}[\psi(X)]. \tag{5.10}$$

Then ρ is proper, convex lsc functional and, for every $X \in \operatorname{dom}(\rho)$,

$$\partial_X \rho(X) = \{Q \in \mathcal{L}^2 : Q \in \partial\psi(X)\ \mathbb{P}_0 - a.s.\}. \tag{5.11}$$

Proposition 5.1. *Denote by C^* and t^* the optimal C and t in the primal representation (4.1) of extended φ-divergence risk measure. The risk identifier of risk measure $\mathcal{R}_{\varphi,\beta}(X)$ can be expressed as follows*

$$Q^*(\omega) \in \partial\varphi^* \left(\frac{X(\omega)}{t^*} - C^* \right). \tag{5.12}$$

Denote by C^* the optimal C in the primal representation (4.4) of extended φ-divergence risk measure. The risk identifier of extended φ-divergence error measure $\mathcal{E}_{\varphi,\beta}(X)$ can be expressed as follows

$$Q^*(\omega) \in \partial\varphi^* \left(\frac{X(\omega)}{t^*} \right). \tag{5.13}$$

Proof. It is known that the risk identifier is the subgradient of the risk function (see, for example, Proposition 8.36 of Royset and Wets [2022]). Therefore, (5.12) is obtained by taking the subdifferential of (4.1) following Lemma 5.2. The expression (5.13) is obtained analogously. □

Note that the envelope $\mathcal{Q}_{\varphi,\beta}$ of error does not have the constraint $\mathbb{E}Q = 1$. However, when we minimize $\mathcal{E}_{\varphi,\beta}(X - C)$ with respect to C to get statistic $\mathcal{S}_{\varphi,\beta}(X)$, the constraint $\mathbb{E}Q = 1$ is satisfied automatically. This can be seen from the necessary condition for saddle point (C^*, Q^*)

$$\frac{\partial}{\partial C}\mathbb{E}[(X - C)(Q^* - 1)]\Big|_{C=C^*} = 0. \tag{5.14}$$

6 Relation to Generalized Entropic Risk Measure

This section discusses the relation between the extended φ-divergence risk measure and generalized entropic risk measure.

Proposition 6.1. *The following relation holds for φ-divergence risk measure $\mathcal{R}_{\varphi,\beta}(X)$ and generalized entropic risk measure $\mathcal{R}_\varphi(X)$*

$$\mathcal{R}_{\varphi,\beta}(X) = \inf_{t>0} \{t\beta + \mathcal{R}_{t\varphi}(X)\}. \tag{6.1}$$

Proof. Notice that

$$\begin{aligned}
\mathcal{R}_{t\varphi}(X) =& \inf_C \{C + \mathbb{E}[(t\varphi)^*(X - C)]\} && (6.2)\\
=& \inf_C \left\{ C + t\mathbb{E}\left[\varphi^*\left(\frac{X - C}{t}\right)\right]\right\} && (6.3)\\
=& t \inf_C \left\{ \frac{C}{t} + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - \frac{C}{t}\right)\right]\right\} && (6.4)\\
=& t\mathcal{R}_\varphi\left(\frac{X}{t}\right). && (6.5)
\end{aligned}$$

Then, we have

$$\mathcal{R}_{\varphi,\beta}(X) = \inf_{\substack{C\in\mathbb{R},\\ t>0}} t\Bigg\{\beta + \Big(C + \mathbb{E}\Big[\varphi^*\Big(\frac{X}{t} - C\Big)\Big]\Big)\Bigg\} \tag{6.6}$$

$$= \inf_{t>0} t\left\{\beta + \mathcal{R}_{\varphi}\left(\frac{X}{t}\right)\right\} \tag{6.7}$$

$$= \inf_{t>0} \left\{t\beta + \mathcal{R}_{t\varphi}\left(X\right)\right\}. \tag{6.8}$$

Equality (6.7) is obtained by (2.12) and equality (6.8) is obtained by (6.5). □

We recover a special case of Theorem 4.1 of Rockafellar [2023] by directly observing the relation in primal representation. Rockafellar [2023] proves the relation for general divergence by studying dual representation and viewing $\mathcal{R}_{\varphi,\beta}(X)$ as the support function of the level set $\mathcal{Q}^{\mathcal{R}}_{\varphi,\beta}$.

Next, we recover the φ-divergence from the elements in the corresponding φ-divergence quadrangle.

Proposition 6.2. *Let $\varphi(x)$ be a divergence function. φ-divergence can be recovered from the elements in φ-divergence quadrangle by*

$$D_{\varphi}(P||P_0) = \sup_{X\in\mathcal{L}^2,\beta>0} \{\mathbb{E}[XQ] - \mathcal{R}_{\varphi,\beta}\left(X\right) - \beta\} \tag{6.9}$$

$$= \sup_{X\in\mathcal{L}^2,\beta>0} \{\mathbb{E}[X(Q-1)] - \mathcal{D}_{\varphi,\beta}\left(X\right) - \beta\} \tag{6.10}$$

$$= \sup_{X\in\mathcal{L}^2,\beta>0,C} \{\mathbb{E}[XQ] - \mathcal{V}_{\varphi,\beta}\left(X-C\right) + C - \beta\} \tag{6.11}$$

$$= \sup_{X\in\mathcal{L}^2,\beta>0,C} \{\mathbb{E}[X(Q-1)] - \mathcal{E}_{\varphi,\beta}\left(X-C\right) - \beta\}. \tag{6.12}$$

Proof. From (6.8), we have by convex conjugate

$$\mathcal{R}_{t\varphi}\left(X\right) = \inf_{\beta>0}\{t\beta + \mathcal{R}_{\varphi,\beta}\left(X\right)\}. \tag{6.13}$$

(6.13) is a generalization of Proposition 3.1 in Föllmer and Knispel [2011].

Next, we have

$$\mathbb{E}[\varphi(Q)] = \sup_{X\in\mathcal{L}^2}\{\mathbb{E}[XQ] - \mathcal{R}_{\varphi}(X)\} \tag{6.14}$$

$$= \sup_{X\in\mathcal{L}^2}\{\mathbb{E}[XQ] - \inf_{\beta>0}\{\beta + \mathcal{R}_{\varphi,\beta}\left(X\right)\}\} \tag{6.15}$$

$$= \sup_{X\in\mathcal{L}^2,\beta>0}\{\mathbb{E}[XQ] - \mathcal{R}_{\varphi,\beta}\left(X\right) - \beta\}, \tag{6.16}$$

where (6.14) is by (2.11), (6.15) is by plugging in (6.13) to (2.11).

Since $\varphi(x)$ is a divergence function, $\mathbb{E}[\varphi(Q)] = D_{\varphi}(P||P_0)$. The rest of the proof is by quadrangle relations. □

Discussion on Risk Quadrangle The primal representation (2.12) has the form of expectation-type risk measure [Rockafellar and Uryasev, 2013]. However, it is not regular or coherent in general.

Still, we can mimic the quadrangle elements and define the following elements

$$\mathcal{R}_\varphi(X) = \inf_{C\in\mathbb{R}} \Big\{C + \mathbb{E}[\varphi^*(X-C)]\Big\} \tag{6.17}$$
$$\mathcal{D}_\varphi(X) = \inf_{C\in\mathbb{R}} \Big\{C + \mathbb{E}[\varphi^*(X-C) - X]\Big\} \tag{6.18}$$
$$\mathcal{V}_\varphi(X) = \mathbb{E}[\varphi^*(X)] \tag{6.19}$$
$$\mathcal{E}_\varphi(X) = \mathbb{E}[\varphi^*(X) - X] \tag{6.20}$$
$$\mathcal{S}_\varphi(X) = \operatorname*{argmin}_{C\in\mathbb{R}} \Big\{C + \mathbb{E}[\varphi^*(X-C)]\Big\}. \tag{6.21}$$

Quadrangle relations in Definition 2.5 are satisfied.

Since $(t\varphi)^*(x) = t\varphi^*\left(X/t\right)$,

$$\mathcal{R}_{t\varphi}(X) = \inf_{C\in\mathbb{R}} t\left\{C + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - C\right)\right]\right\} \tag{6.22}$$
$$\mathcal{D}_{t\varphi}(X) = \inf_{C\in\mathbb{R}} t\left\{C + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - C\right) - \frac{X}{t}\right]\right\} \tag{6.23}$$
$$\mathcal{V}_{t\varphi}(X) = t\mathbb{E}\left[\varphi^*\left(\frac{X}{t}\right)\right] \tag{6.24}$$
$$\mathcal{E}_{t\varphi}(X) = t\mathbb{E}\left[\varphi^*\left(\frac{X}{t}\right) - \frac{X}{t}\right] \tag{6.25}$$
$$\mathcal{S}_{t\varphi}(X) = \operatorname*{argmin}_{C\in\mathbb{R}} t\left\{C + \mathbb{E}\left[\varphi^*\left(\frac{X}{t} - C\right)\right]\right\}. \tag{6.26}$$

Plugging in to Theorem 4.1, we can express the elements in the extended φ-divergence quadrangle with (6.17)-(6.21)

$$\mathcal{R}_{\varphi,\beta}(X) = \inf_{t>0} \left\{t\beta + \mathcal{R}_{t\varphi}\left(X\right)\right\} \tag{6.27}$$
$$\mathcal{D}_{\varphi,\beta}(X) = \inf_{t>0} \left\{t\beta + \mathcal{D}_{t\varphi}\left(X\right)\right\} \tag{6.28}$$
$$\mathcal{V}_{\varphi,\beta}(X) = \inf_{t>0} \left\{t\beta + \mathcal{V}_{t\varphi}\left(X\right)\right\} \tag{6.29}$$
$$\mathcal{E}_{\varphi,\beta}(X) = \inf_{t>0} \left\{t\beta + \mathcal{E}_{t\varphi}\left(X\right)\right\} \tag{6.30}$$
$$\mathcal{S}_{\varphi,\beta}(X) = \operatorname*{argmin}_{C\in\mathbb{R}} \inf_{t>0} \left\{C + t\beta + \mathcal{R}_{t\varphi}\left(X - C\right)\right\}. \tag{6.31}$$

Rockafellar [2023] proposes a parent quadrangle and studies its relation to a more general version of the divergence quadrangle. (6.17)–(6.21) is a special case of parent quadrangle. (6.27)–(6.31) is a special case of the relations in Rockafellar [2023]. Rockafellar [2023] also studies subaverse quadrangle that includes the parent quadrangle in the Fundamental Risk Quadrangle framework. The difference in this discussion is that we observe the relation directly from the primal representation.

7 Examples

This section lists some examples of extended φ-divergence and φ-divergence quadrangles. Risk measures in many of the considered examples are studied in existing literature. We provided the complete quadrangles and, consequently, made the connections to corresponding regressions.

7.1 Examples of Extended φ-Divergence Quadrangles

Example 1 (Range-based Quadrangle Generated by Extended Total Variation Distance)**.** Consider the following extended divergence function and its convex conjugate

$$\varphi(x) = |x-1|, \quad x \in \mathbb{R} \text{ and } \varphi^*(z) = \begin{cases} z, & z \in [-1,1] \\ +\infty, & z \in (-\infty,-1) \cup (1,+\infty) \ . \end{cases} \tag{7.1}$$

The quadrangle elements are infinite if X is unbounded. Thus we consider $X \in \mathcal{L}^\infty$ to obtain nontrivial expression. The regret measure is given by

$$\begin{aligned} \mathcal{V}_{\varphi,\beta}(X) &= \inf_{\substack{t>0 \\ t\geq -\operatorname{ess\,inf} X \\ t\geq \operatorname{ess\,sup} X}} \{t\beta + \mathbb{E}[X]\} && (7.2)\\ &= \beta \max\{0, -\operatorname{ess\,inf} X, \operatorname{ess\,sup} X\} + \mathbb{E}[X] && (7.3)\\ &= \beta \operatorname{ess\,sup} |X| + \mathbb{E}[X] \ . && (7.4) \end{aligned}$$

The risk measure is given by

$$\begin{aligned} \mathcal{R}_{\varphi,\beta}(X) &= \inf_{\substack{t>0,\, C\in\mathbb{R} \\ t(C-1)\leq \operatorname{ess\,inf} X \\ t(C+1)\geq \operatorname{ess\,sup} X}} \{t\beta + tC + \mathbb{E}[X - tC]\} && (7.5)\\ &= \frac{\beta}{2}(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X) + \mathbb{E}[X] \ . && (7.6) \end{aligned}$$

From the constraints $t(C-1) \leq \operatorname{ess\,inf} X$ and $t(C+1) \geq \operatorname{ess\,sup} X$, we have

$$2t \geq \operatorname{ess\,sup} X - \operatorname{ess\,inf} X,$$

hence the optimal

$$t^* = (\operatorname{ess\,sup} X - \operatorname{ess\,inf} X)/2.$$

From the constraints, we have

$$2\operatorname{ess\,sup} X/(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X) - 1 \leq C \leq 2\operatorname{ess\,inf} X/(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X) + 1.$$

Thus,

$$(\operatorname{ess\,sup} X + \operatorname{ess\,inf} X)/(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X) \geq C \geq (\operatorname{ess\,sup} X + \operatorname{ess\,inf} X)/(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X),$$

yelding

$$C^* = (\operatorname{ess\,sup} X + \operatorname{ess\,inf} X)/(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X).$$

Therefore, the statistic

$$\mathcal{S}_{\varphi,\beta} = C^* t^* = (\operatorname{ess\,sup} X + \operatorname{ess\,inf} X)/2.$$

The complete quadrangle is as follows

$$\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= \frac{\beta}{2}(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X) + \mathbb{E}[X] = \text{ range-buffered risk, scaled} \\
\mathcal{V}_{\varphi,\beta}(X) &= \beta \operatorname{ess\,sup} |X| + \mathbb{E}[X] = \text{ } \mathcal{L}^{\infty}\text{-regret, scaled} \\
\mathcal{D}_{\varphi,\beta}(X) &= \frac{\beta}{2}(\operatorname{ess\,sup} X - \operatorname{ess\,inf} X) = \text{ radius of the range, scaled} \\
\mathcal{E}_{\varphi,\beta}(X) &= \beta \operatorname{ess\,sup} |X| = \text{ } \mathcal{L}^{\infty}\text{-error, scaled} \\
\mathcal{S}_{\varphi,\beta}(X) &= \frac{1}{2}(\operatorname{ess\,sup} X + \operatorname{ess\,inf} X) = \text{ center of range, scaled}
\end{aligned}$$

We recovered the range-based quadrangle in Example 4 of Rockafellar and Uryasev [2013]. The divergence function is the extended version of the divergence function of total variation distance in Example 6.

Example 2 (Mean Quadrangle Generated by Extended Pearson χ^2-divergence)**.** Consider the following extended divergence function and its convex conjugate

$$\varphi(x) = (x-1)^2 \quad \text{and} \quad \varphi^*(z) = \frac{z^2}{4} + z\,. \tag{7.7}$$

Then, the extended Pearson χ^2-divergence risk measure is given by

$$\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= \inf_{t>0, C\in\mathbb{R}} t\left\{C + \beta + \frac{1}{4t^2}\mathbb{E}[(X-C)^2] + \mathbb{E}[\frac{X-C}{t}]\right\} \\
&= \inf_{t>0, C\in\mathbb{R}} \left\{t\beta + \frac{1}{4t}\mathbb{E}[(X-C)^2] + \mathbb{E}[X]\right\} \\
&= \mathbb{E}[X] + \sqrt{\beta\mathbb{V}[X]},
\end{aligned}$$

where $\mathbb{V}[X] = \mathbb{E}[(X-\mathbb{E}[X])^2]$ is the variance of X and (t^*, C^*), which furnish the minimum are

$$t^* = \sqrt{\frac{\mathbb{V}[X]}{4\beta}}, \qquad C^* = \mathbb{E}[X].$$

Evidently, the corresponding regret is given by

$$\begin{aligned}
\mathcal{V}_{\varphi,\beta}(X) &= \mathbb{E}[X] + \sqrt{\beta\mathbb{E}[X^2]} \\
&= \mathbb{E}[X] + \sqrt{\beta}\,\|X\|_2\,.
\end{aligned}$$

Let $\lambda = \sqrt{\beta}$ and $\sqrt{\mathbb{V}[X]} = \sigma(X)$, then the complete quadrangle is as follows

$$\begin{aligned}
\mathcal{R}_{\varphi,\lambda}(X) &= \mathbb{E}[X] + \lambda\sigma(X) = \text{ safety margin tail risk} \\
\mathcal{V}_{\varphi,\lambda}(X) &= \mathbb{E}[X] + \lambda\,\|X\|_2 = \text{ } \mathcal{L}^2\text{-regret, scaled} \\
\mathcal{D}_{\varphi,\lambda}(X) &= \lambda\sigma(X) = \text{ standard deviation, scaled} \\
\mathcal{E}_{\varphi,\lambda}(X) &= \lambda\,\|X\|_2 = \text{ } \mathcal{L}^2\text{-error, scaled} \\
\mathcal{S}_{\varphi,\lambda}(X) &= \mathbb{E}[X] = \text{ mean}
\end{aligned}$$

We recovered the mean quadrangle in Example 1 of Rockafellar and Uryasev [2013]. The divergence function is the extended version of the divergence function of χ^2-divergence in Example 7. It is worth noting that the radius β of the uncertainty set does not impact the regression result, since it only impact the scale the error function.

Example 3 (Expectile Quadrangle Generated by Generalized Pearson χ^2-divergence)**.** Let $0 < p < 1$. Consider the following extended divergence function and its convex conjugate

$$\varphi(x) = \begin{cases} \frac{1}{q}(x-1)^2, & x > 1 \\ \frac{1}{1-q}(x-1)^2, & x \le 1 \end{cases} \text{ and } \varphi^*(z) = \begin{cases} (\frac{qz}{2}+1)z - \frac{1}{q}(\frac{qz}{2})^2 = \frac{qz^2}{4} + z, & z > 0 \\ (\frac{(1-q)z}{2}+1)z - \frac{1}{1-q}(\frac{(1-q)z}{2})^2 = \frac{(1-q)z^2}{4} + z, & z \le 0 \end{cases}. \tag{7.8}$$

The error measure is given by

$$\mathcal{E}_{\varphi,\beta}(X) = \inf_{t>0} t\beta + \mathbb{E}\left[t\varphi^*\left(\frac{X}{t}\right) - X\right] \tag{7.9}$$

$$= \inf_{t>0} t\beta + \frac{1}{4t}E\left[qX_-^2 + (1-q)X_+^2\right] \tag{7.10}$$

$$= t\beta + \frac{1}{4t}E\left[qX_-^2 + (1-q)X_+^2\right]\Bigg|_{t=\sqrt{\frac{E\left[qX_-^2+(1-q)X_+^2\right]}{4\beta}}} \tag{7.11}$$

$$= \sqrt{\beta\mathbb{E}\left[qX_-^2 + (1-q)X_+^2\right]} \tag{7.12}$$

The complete quadrangle is as follows

$$\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= q\mathbb{E}[(((X - e_q(X))_+)^2] + (1-q)\mathbb{E}[(((X - e_q(X))_-)^2] + \mathbb{E}[X] \\
\mathcal{V}_{\varphi,\beta}(X) &= \mathbb{E}[X] + \sqrt{\beta E\left[qX_-^2 + (1-q)X_+^2\right]} \\
\mathcal{D}_{\varphi,\beta}(X) &= q\mathbb{E}[(((X - e_q(X))_+)^2] + (1-q)\mathbb{E}[(((X - e_q(X))_-)^2] \\
\mathcal{E}_{\varphi,\beta}(X) &= \sqrt{\beta E\left[qX_-^2 + (1-q)X_+^2\right]} = \text{ asymmetric squared loss, scaled} \\
\mathcal{S}_{\varphi,\beta}(X) &= e_q(X) = \text{ expectile}
\end{aligned}$$

We recover one version of expectile quadrangle in Malandii et al. [2024]. The divergence function $\varphi(x)$ gives rise to a generalized Pearson χ^2-divergence. Example 2 is a special case of this quadrangle with $q = 0.5$.

7.2 Examples: φ-Divergence Quadrangle

Example 4 (EVaR Quadrangle Generated by Kullback-Leibler Divergence)**.** The divergence function and its convex conjugate are

$$\varphi(x) = x\ln(x) - x + 1, \quad \varphi^*(z) = \exp(z) - 1.$$

Let $\beta = \ln\left(\frac{1}{1-\alpha}\right)$. The complete quadrangle is as follows:

$$\mathcal{R}_{\varphi,\alpha}(X) = \mathrm{EVaR}_\alpha(X) = \inf_{t>0} t\left\{ \ln \mathbb{E}\Big[\frac{e^{\frac{X}{t}}}{1-\alpha}\Big] \right\},$$

$$\mathcal{D}_{\varphi,\alpha}(X) = \mathrm{EVaR}_\alpha(X) - \mathbb{E}[X] = \inf_{t>0} t\left\{ \ln \mathbb{E}\Big[\frac{e^{\frac{X-\mathbb{E}[X]}{t}}}{1-\alpha}\Big] \right\},$$

$$\mathcal{V}_{\varphi,\alpha}(X) = \inf_{t>0} t\left\{ \ln\Big(\frac{1}{1-\alpha}\Big) + \mathbb{E}\big[e^{\frac{X}{t}} - 1\big] \right\},$$

$$\mathcal{E}_{\varphi,\alpha}(X) = \inf_{t>0} t\left\{ \ln\Big(\frac{1}{1-\alpha}\Big) + \mathbb{E}\Big[e^{\frac{X}{t}} - \frac{X}{t} - 1\Big] \right\},$$

$$\mathcal{S}_{\varphi,\alpha}(X) = t^* \ln \mathbb{E}\big[e^{\frac{X}{t^*}}\big],$$

In the quadrangle, t^* is a solution of the following equation:

$$t^* \ln\Big(\frac{1}{1-\alpha}\Big) + t^* \ln \mathbb{E}\big[e^{\frac{X}{t^*}}\big] - \frac{\mathbb{E}\big[Xe^{\frac{X}{t^*}}\big]}{\mathbb{E}\big[e^{\frac{X}{t^*}}\big]} = 0,$$

which can be obtained from the second equation in (5.9) when $\varphi^*(z) = \exp(z) - 1$. The risk measure in this quadrangle is studied in Ahmadi-Javid [2012].

Example 5 (Quantile Quadrangle Generated by Indicator Divergence)**.** Consider the divergence function and its convex conjugate

$$\varphi(x) = \mathbf{1}_{[0,(1-\alpha)^{-1}]}(x), \qquad \varphi^*(z) = \max\{0, (1-\alpha)^{-1} z\}.$$

We obtain the Quantile Quadrangle:

$$\mathcal{R}_{\varphi,\beta}(X) = \inf_{\substack{C\in\mathbb{R},\\ t>0}} t\left\{ C + \beta + \frac{1}{1-\alpha}\mathbb{E}\Big[\frac{X}{t} - C\Big]_+ \right\} = \mathrm{CVaR}_\alpha(X), \tag{7.13}$$

$$\mathcal{V}_{\varphi,\beta}(X) = \inf_{t>0} t\left\{ \beta + \frac{1}{1-\alpha}\mathbb{E}\Big[\frac{X}{t}\Big]_+ \right\} = \frac{1}{1-\alpha}\mathbb{E}[X_+], \tag{7.14}$$

$$\mathcal{D}_{\varphi,\beta}(X) = \inf_{\substack{C\in\mathbb{R},\\ t>0}} t\left\{ C + \beta + \frac{1}{\alpha}\mathbb{E}\Big[\Big[\frac{X}{t} - C\Big]_+ - \frac{X}{t}\Big] \right\} = \mathrm{CVaR}_\alpha(X) - \mathbb{E}[X], \tag{7.15}$$

$$\mathcal{E}_{\varphi,\beta}(X) = \inf_{t>0} t\left\{ \beta + \frac{1}{1-\alpha}\mathbb{E}\Big[\Big[\frac{X}{t}\Big]_+ - \frac{X}{t}\Big] \right\} = \mathbb{E}\Big[\frac{\alpha}{1-\alpha}X_+ + X_-\Big], \tag{7.16}$$

$$\mathcal{S}_{\varphi,\beta}(X) = \operatorname*{argmin}_{C\in\mathbb{R}} \inf_{t>0} t\left\{ \frac{C}{t} + \beta + \frac{1}{1-\alpha}\mathbb{E}\Big[\frac{X-C}{t}\Big]_+ \right\} = \mathrm{VaR}_\alpha(X). \tag{7.17}$$

The derivation of the risk measure (7.13) is from Ahmadi-Javid [2012]; Shapiro [2017]. We recover the quantile quadrangle in Example 2 of Rockafellar and Uryasev [2013]. Note that the radius β of the divergence ball does not appear in the formula in the primal representation. When $\alpha \to 1$, the quadrangle becomes the worst-case-based quadrangle. When $\alpha \to 0$, the risk measure becomes $\mathbb{E}[X]$, which is not risk averse. $\varphi(x)$ in this case violates Definition 2.9.

Example 6 (Robustified Supremum-Based Quadrangle Generated by Total Variation Distance)**.** Consider the following divergence function and its convex conjugate

$$\varphi(x) = \begin{cases} |x-1|, & x \geq 0 \\ +\infty, & x < 0 \end{cases} \text{ and } \varphi^*(z) = \begin{cases} -1 + [z+1]_+, & z \leq 1 \\ +\infty, & z > 1 \end{cases}. \tag{7.18}$$

The risk measure is given by

$$\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= \inf_{\substack{t>0,\, C\in\mathbb{R} \\ \operatorname{ess\,sup}(X-C)\leq t}} \{t\beta + C - t + \mathbb{E}[X - C + t]_+\} \\
&= \inf_{\substack{t>0,\, C\in\mathbb{R} \\ \operatorname{ess\,sup}(X-C-t)\leq t}} \{t\beta + C + \mathbb{E}[X - C]_+\} \\
&= \inf_{\substack{t>0,\, C\in\mathbb{R} \\ \operatorname{ess\,sup}(X)-2t\leq C}} \{t\beta + C + \mathbb{E}[X - C]_+\}.
\end{aligned}$$

The function being minimized is convex in C. It attains minimum at $C \in (-\infty, \operatorname{ess\,inf} X]$ if there is no constraint on C. Thus the minimum in C is attained at $C^* = \operatorname{ess\,sup}(X) - 2t$. Suppose that $\beta \in (0,2)$ (Note that TVD is no larger than 2). Then

$$\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= \operatorname{ess\,sup}(X) + \inf_{t>0} \{t(\beta - 2) + \mathbb{E}[X - \operatorname{ess\,sup}(X) + 2t]_+\} \\
&= \operatorname{ess\,sup}(X) + \inf_{t<0} \left\{ t(1 - \frac{\beta}{2}) + \mathbb{E}[X - \operatorname{ess\,sup}(X) - t]_+ \right\} \\
&= \operatorname{ess\,sup}(X) + (1 - \frac{\beta}{2}) \inf_{t<0} \left\{ t + (1 - \frac{\beta}{2})^{-1} \mathbb{E}[X - \operatorname{ess\,sup}(X) - t]_+ \right\}.
\end{aligned}$$

Note that since $X - \operatorname{ess\,sup}(X) \leq 0$, the minimum in the last equation is attained at some $t \leq 0$, and this minimum is equal to

$$\mathrm{CVaR}_{\frac{\beta}{2}}(X - \operatorname{ess\,sup}(X)) = \mathrm{CVaR}_{\frac{\beta}{2}}(X) - \operatorname{ess\,sup}(X).$$

Therefore, the complete quadrangle is as follows:

$$\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= \frac{\beta}{2} \operatorname{ess\,sup}(X) + (1 - \frac{\beta}{2}) \mathrm{CVaR}_{\frac{\beta}{2}}(X), \\
\mathcal{V}_{\varphi,\beta}(X) &= \inf_{t>0, t\geq \operatorname{ess\,sup} X} \left\{ t(\beta - 1) + \mathbb{E}\Big[X + t\Big]_+ \right\}, \\
\mathcal{D}_{\varphi,\beta}(X) &= \frac{\beta}{2} \operatorname{ess\,sup}(X) + (1 - \frac{\beta}{2}) \mathrm{CVaR}_{\frac{\beta}{2}}(X) - \mathbb{E}[X], \\
\mathcal{E}_{\varphi,\beta}(X) &= \inf_{t>0} \left\{ t(\beta - 1) + \mathbb{E}\Big[\Big[X + t\Big]_+ - X\Big] \right\}, \\
\mathcal{S}_{\varphi,\beta}(X) &= \operatorname{ess\,sup}(X) - 2\mathrm{VaR}_{1-\frac{\beta}{2}}(X).
\end{aligned}$$

The derivation of the risk measure is studied in Example 3.10 of Shapiro [2017]. From the dual representation of the risk measure, one can intuitively understand the shape of the worst-case distribution. Since the total variation distance is L_1 distance, the worst-case distribution is obtained by moving the probability of the smallest element to the largest one, while keeping the others intact. If the radius is sufficiently large, then after moving all the the probability of the smallest element, the probability of the second smallest element will be moved to that of the largest element, and so on.

Since the risk envelope corresponding to $\varphi(x)$ in (7.18) is a subset of the risk envelope generated by the extended version in Example 1, the risk, regret, deviation and error in this quadrangle is an lower bound of those in Example 1.

Example 7 (Second-order Quantile-based Quadrangle Generated by Pearson χ^2-divergence)**.** The divergence function and its convex conjugate are

$$\varphi(x)=\begin{cases}x^2-1, & x\geq 0\\ +\infty, & x<0\end{cases} \quad \text{and} \quad \varphi^*(z)=\begin{cases}\frac{z^2}{4}+1, & z\geq 0\\ 1, & z<0\end{cases}=1+\frac{z^2}{4}I_{z\geq 0}\,. \tag{7.19}$$

We derive the corresponding error measure

$$\mathcal{E}_{\varphi,\beta}(X)=\inf_{t>0} t\left\{\beta+\mathbb{E}\left[\varphi^*\left(\frac{X}{t}\right)-\frac{X}{t}\right]\right\} \tag{7.20}$$

$$=\inf_{t>0}\left\{t(\beta+1)+\frac{1}{4t}\mathbb{E}\left[X^2 I_{\{X\geq 0\}}\right]-\mathbb{E}[X]\right\} \tag{7.21}$$

$$=\sqrt{(\beta+1)\mathbb{E}\left[X^2 I_{\{X\geq 0\}}\right]}-\mathbb{E}[X]\,. \tag{7.22}$$

The complete quadrangle is

$$\mathcal{R}_{\varphi,\beta}(X)=\min_{C\in\mathbb{R}}\sqrt{(\beta+1)\mathbb{E}\left[(X-C)^2 I_{\{X\geq C\}}\right]}+C= \text{ second-order superquantile,}$$

$$\mathcal{V}_{\varphi,\beta}(X)=\sqrt{(\beta+1)\mathbb{E}\left[X^2 I_{\{X\geq 0\}}\right]}= \text{ 2-normed absolute loss, scaled,}$$

$$\mathcal{D}_{\varphi,\beta}(X)=\min_{C\in\mathbb{R}}\sqrt{(\beta+1)\mathbb{E}\left[(X-C)^2 I_{\{X\geq C\}}\right]}-\mathbb{E}[X-C]= \text{ second-order superquantile deviation,}$$

$$\mathcal{E}_{\varphi,\beta}(X)=\sqrt{(\beta+1)\mathbb{E}\left[X^2 I_{\{X\geq 0\}}\right]}-\mathbb{E}[X]= \text{ second-order quantile error,}$$

$$\mathcal{S}_{\varphi,\beta}(X)=\operatorname*{argmin}_{C\in\mathbb{R}}\sqrt{(\beta+1)\mathbb{E}\left[(X-C)^2 I_{\{X\geq C\}}\right]}-\mathbb{E}[X-C]= \text{ second-order quantile.}$$

The risk measure is studied in Krokhmal [2007] . This quadrangle a special case of the higher-order quantile-based quadrangle in Example 12 of Rockafellar and Uryasev [2013].

Observe that $\mathcal{V}_{\varphi,\beta}(X-\mathbb{E}[X])=\mathcal{E}_{\varphi,\beta}(X-\mathbb{E}[X])$ is the semideviation [Markowitz, 1959], a popular measure of downside risk.

Since the risk envelope corresponding to $\varphi(x)$ in (7.19) is a subset of the risk envelope generated by the extended version in Example 2, the risk, regret, deviation and error in this quadrangle is an lower bound of those in Example 2.

Example 8 (Example Generated by Two-sided-indicator Divergence)**.** Let $0<a<1<b$. The divergence function and its convex conjugate are

$$\varphi(x)=\begin{cases}+\infty, & x\in[0,a)\\ 0, & x\in[a,b]\\ +\infty, & x\in(b,+\infty)\end{cases}\,, \quad \varphi^*(z)=\begin{cases}az, & z<0\\ bz, & z\geq 0\end{cases}\,. \tag{7.23}$$

The error measure is

$$\mathcal{E}_{\varphi,\beta}(X) = \mathbb{E}[(1-a)X_- + (b-1)X_+] \ . \tag{7.24}$$

The complete quadrangle is

$$\begin{aligned}
\mathcal{R}_{\varphi,\beta}(X) &= (1-a)\mathrm{CVaR}_{\frac{b-1}{b-a}}(X) + a\mathbb{E}[X],\\
\mathcal{V}_{\varphi,\beta}(X) &= \mathbb{E}[(2-a)X_- + bX_+],\\
\mathcal{D}_{\varphi,\beta}(X) &= (1-a)\mathrm{CVaR}_{\frac{b-1}{b-a}}(X) + (a-1)\mathbb{E}[X],\\
\mathcal{E}_{\varphi,\beta}(X) &= \mathbb{E}[(1-a)X_- + (b-1)X_+],\\
\mathcal{S}_{\varphi,\beta}(X) &= \operatorname*{argmin}_{C\in\mathbb{R}} \mathbb{E}[(1-a)(X-C)_- + (b-1)(X-C)_+],
\end{aligned}$$

The risk measure in this quadrangle is studied in Pflug and Ruszczynski [2004], in Ben-Tal and Teboulle [2007] (see Example 2.3), in Love and Bayraksan [2015] (see Example 3). CVaR is a special case of this risk measure for $a = 0$. When $\alpha/(1-\alpha) = (b-1)/(1-a)$, the quadrangle is a scaled version of Example 5.

The risk measure provides another way to connect expectile $e_q(X)$ with distributionally robust optimization (see Proposition 9 in Bellini et al. [2014])

$$e_q(X) = \max_{\gamma\in[\frac{1-q}{q},1]} \mathcal{R}_{I_{[\gamma,\gamma\frac{q}{1-q}]},\beta}(X) \ . \tag{7.25}$$

8 Robust Optimization Interpretation for Various Applications

This section generalizes robust optimization interpretations presented in the Introduction. The primal representations of the risk measures and error measures in Section 4 are frequently used as an objective function in optimization and regression. From the dual representation in Section 3, we obtain the general interpretation as robust optimization. In particular, examples in Section 7 all have such interpretation.

We start with the interpretation of risk minimization.

Risk Minimization as a Robust Expected Loss Minimization Consider an extended divergence function $\varphi(x)$ and the corresponding risk measure $\mathcal{R}_{\varphi,\beta}(X)$ in the quadrangle in Theorem 3.1. Denote by $\mathcal{X}$ the set of feasible random variables. Risk minimization (8.1) can be interpreted as robust loss minimization (8.2)

Risk minimization	**Robust loss minimization**
$\min_{X\in\mathcal{X}} \mathcal{R}_{\varphi,\beta}(X)\ , \quad (8.1)$	$\min_{X\in\mathcal{X}} \max_{Q\in\mathcal{Q}_{\varphi,\beta}} \mathbb{E}[QX]\ . \quad (8.2)$

Problem (8.2) is the robust version of the expected loss minimization problem $\min_{X\in\mathcal{X}} \mathbb{E}[X]$. If $\varphi(x)$ is a divergence function defined in Definition 2.9, risk minimization (8.1) can be interpreted as distributionally robust loss minimization (8.3)

Distributionally robust loss minimization

$$\min_{X\in\mathcal{X}} \max_{P\in\mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[X] \tag{8.3}$$

Portfolio Optimization as a Robust Expected Loss Minimization Following the setup above, we consider the portfolio loss $X(\boldsymbol{w})$, where $\boldsymbol{w}$ is the portfolio weight. Portfolio optimization (8.4) can be interpreted as robust loss minimization (8.5)

Portfolio Optimization	**Robust loss minimization**
$$\min_{\boldsymbol{w}:\mathbf{1}^T\boldsymbol{w}=1} \mathcal{R}_{\varphi,\beta}(X(\boldsymbol{w})) \ , \qquad (8.4)$$	$$\min_{\boldsymbol{w}:\mathbf{1}^T\boldsymbol{w}=1} \max_{Q\in\mathcal{Q}_{\varphi,\beta}} \mathbb{E}[QX(\boldsymbol{w})] \ . \qquad (8.5)$$

Problem (8.5) is the robust version of the expected loss minimization problem $\min\limits_{\mathbf{1}^\top\boldsymbol{w}=1} \mathbb{E}[X(\boldsymbol{w})]$.

If $\varphi(x)$ is a divergence function defined in Definition 2.9, portfolio optimization (8.4) can be interpreted as distributionally robust loss minimization (8.6)

Distributionally robust loss minimization

$$\min_{\boldsymbol{w}:\mathbf{1}^T\boldsymbol{w}=1} \max_{P\in\mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[X(\boldsymbol{w})] \tag{8.6}$$

Next, we discuss the interpretation of classification. Consider attribute $\boldsymbol{X}$, label Y and decision vector $\boldsymbol{w}$. The margin is defined by $L(\boldsymbol{w}, b) = Y(\boldsymbol{w}^T\boldsymbol{X} - b)$. $\gamma(\boldsymbol{w})$ is the regularization term.

Classification as Robust Expected Margin Minimization Classification can be interpreted as robust expected margin maximization

Classification	**Robust expected margin maximization**
$$\min_{\boldsymbol{w}} \mathcal{R}_{\varphi,\beta}(-L(\boldsymbol{w}, b)) + \gamma(\boldsymbol{w}) \ , \qquad (8.7)$$	$$\min_{\boldsymbol{w}} \max_{Q\in\mathcal{Q}_{\varphi,\beta}} \mathbb{E}[-QL(\boldsymbol{w}, b)] + \gamma(\boldsymbol{w}) \ . \qquad (8.8)$$

Problem (8.8) is the regularized robust version of the expected margin maximization problem $\max_{\boldsymbol{w}} \mathbb{E}[L(\boldsymbol{w}, b)]$.

If $\varphi(x)$ is a divergence function defined in Definition 2.9, classification (8.7) can be interpreted as distributionally robust loss minimization (8.9)

Distributionally robust expected margin minimization

$$\min_{\boldsymbol{w}:\mathbf{1}^T\boldsymbol{w}=1} \max_{P\in\mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[X(\boldsymbol{w})] + \gamma(\boldsymbol{w}) \ . \tag{8.9}$$

Next, we discuss the interpretation of regression. Consider a dependent variable (regressant) Y, a vector of independent variables (regressors) $\boldsymbol{X} = (X_1, \ldots, X_d)$, a class of function $\mathcal{F}$ and $C \in \mathbb{R}$. The regression residual is defined by $Z_f = Y - f(\boldsymbol{X}) - C$, and the residual without intercept C is defined by $\bar{Z}_f = Y - f(\boldsymbol{X})$.

Regression as a Robust Optimization Consider an extended divergence function $\varphi(x)$ and the corresponding error measure $\mathcal{E}_{\varphi,\beta}(X)$ in the quadrangle in Theorem 3.1. Regression (8.10) can be interpreted as deviation minimization (8.11)(8.12)

Regression	**Deviation minimization**
$$\min_{f\in\mathcal{F},C} \mathcal{E}_{\varphi,\beta}(Z_f)), \qquad (8.10)$$	$$\min_f \left\{ \max_{Q\in\mathcal{Q}_{\varphi,\beta}} \mathbb{E}[Q\bar{Z}_f] - \mathbb{E}[\bar{Z}_f] \right\} \qquad (8.11)$$ $$\text{calculate } C = \mathcal{S}(\bar{Z}_f) \ . \qquad (8.12)$$

If $\varphi(x)$ is a divergence function defined in Definition 2.9,the regression (8.10) can be interpreted as deviation minimization (8.13), (8.14), where the uncertainty set is a φ-divergence ball

Deviation minimization

$$\min_f \left\{ \max_{P\in\mathcal{P}_{\varphi,\beta}} \mathbb{E}_P[\bar{Z}_f] - \mathbb{E}[\bar{Z}_f] \right\} \tag{8.13}$$

$$\text{calculate } C = \mathcal{S}(\bar{Z}_f)\ . \tag{8.14}$$

The equivalence between (8.10) and (8.11), (8.12) is proved as follows. By Theorem 2.1, Problem (8.10) is equivalent to

$$\min_f \mathcal{D}_{\varphi,\beta}(\bar{Z}_f) \tag{8.15}$$

$$\text{calculate } C = \mathcal{S}(\bar{Z}_f)\ . \tag{8.16}$$

The equivalence to Problem (8.11)(8.12) follows from the dual representation of $\mathcal{D}_{\varphi,\beta}(X)$ in (3.3).

9 Case Study: Risk Identifier Visualization

This section contains three case studies visualizing the risk envelope in portfolio optimization (8.5), classification (8.8) and regression (8.11), (8.12). The code is available for download[1]. We focus on the mean quadrangle (Example 2) in this case study. We first solve the problems (8.4), (8.7) and (8.10) in primal representations. With the optimal solutions, we obtain the random variable X in three problems, respectively. By plugging in $\varphi^*(z) = z^2/2+1$ to Proposition 5.1, we obtain the risk identifier $Q^* = (X/t^* - C^*)^2/2 + 1$.

Data The data for portfolio optimization and regression are the same: it is generated by drawing 1,000 samples from a bivariate zero-mean Gaussian distribution. The variance of both random variables is 1, while the covariance is 0.5. The data for classification is generated by two normal distributions with different mean and different covariance matrix. The first has mean $(-0.3, 0)$, while the second has mean $(0.3, 0)$. For both distributions, the variance is 0.05 while the covariance is 0.02. The value of the risk identifier Q^* is represented through the intensity of color. Darker points have larger values.

9.1 Portfolio Optimization

We illustrate the idea with Markowitz portfolio optimization from the mean quadrangle (Example 2). The data points (x, y) represents the loss (negative return) of two assets. We choose $\beta = 100$. The optimal portfolio weight is $(0.509, 0.491)$. The value of the risk identifier Q^* is represented through the intensity of color in Figure 1. Darker color corresponds to a larger value. Larger values are assigned to data points incurring larger loss, i.e., points whose both coordinates are larger.

[1]https://uryasev.ams.stonybrook.edu/index.php/risk-envelope-visualization-for-extended-phi-divergence-quadrangle/

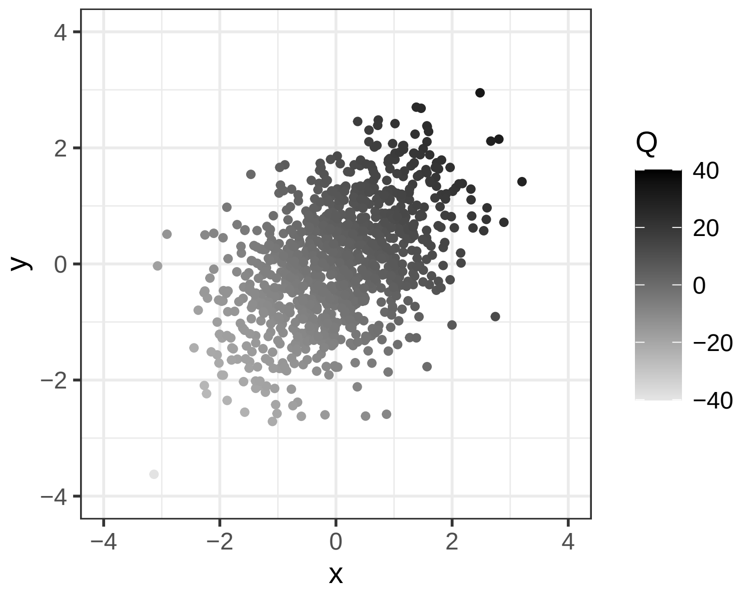

Figure 1: Risk envelope in Markowitz portfolio optimization. Darker points correspond to higher values of $Q^*(w)$. Optimal portfolio weights = $(0.509, 0.491)$.

9.2 Classification

We illustrate the idea with large margin distribution machine from the mean quadrangle (Example 2). We choose $\beta = 0.01$ and $\gamma(\boldsymbol{w}) = ||\boldsymbol{w}||_2^2$. The optimal decision line is $0.129x - 0.009y + 0.002 = 0$. The circles represent samples with label 1, while the diamonds represent samples with label -1. The value of the risk identifier Q^* is represented through the intensity of color in Figure 2. A darker spot corresponds to a larger value. Larger values are assigned to data points incurring larger loss (negative margin), i.e., points that are correctly classified and have larger perpendicular distance from the optimal decision line.

9.3 Regression

We illustrate the idea with least squares regression from the mean quadrangle (Example 2). We choose $\beta = 100$. The regression line is $y = 0.47x + 0.00197$. The value of the risk identifier Q^* is represented through the intensity of color in Figure 3. A darker spot corresponds to a larger value. Larger values are assigned to data points incurring larger loss, i.e., data points further above the regression line.

10 Concluding remarks

We study the primal and dual representation of the extended φ-divergence quadrangle. The quadrangle connects portfolio optimization, classification and regression problem. It includes

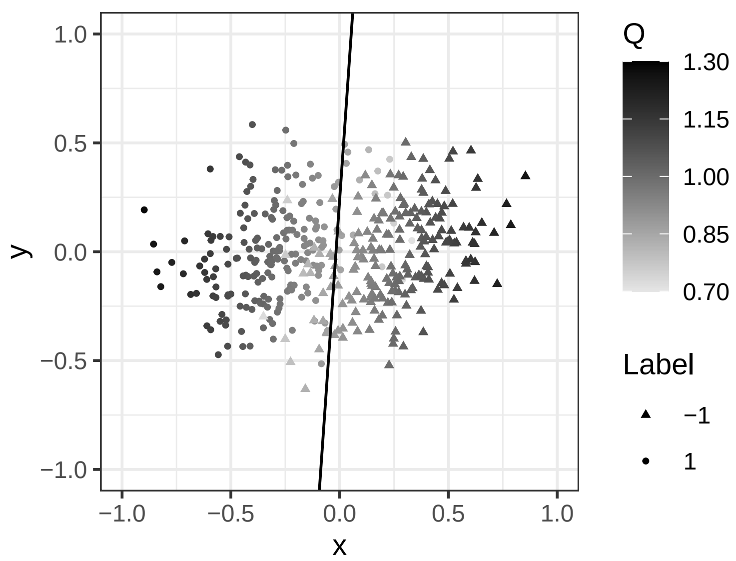

Figure 2: Risk envelope in Large Margin Distribution Machine. Darker points correspond to higher values of $Q^*(w)$. The circles represent samples with label 1, while the diamonds represent samples with label -1. The optimal decision line is $0.129x - 0.009y + 0.002 = 0$.

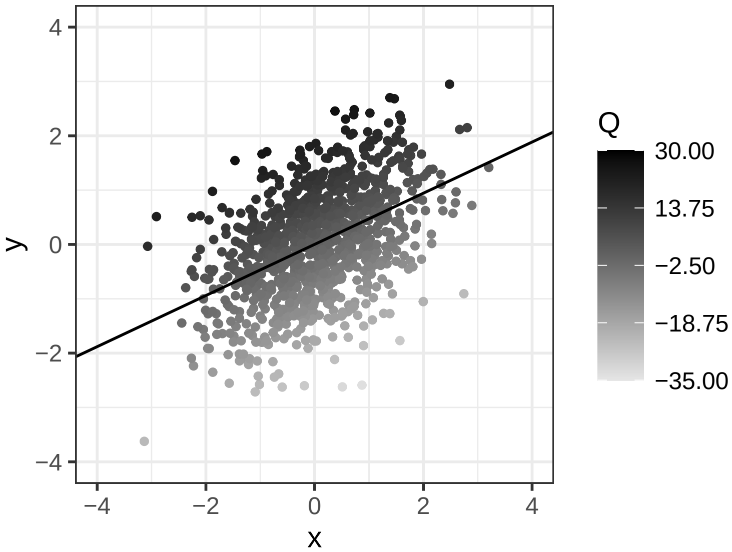

Figure 3: Risk envelope in the least squares regression. Darker points correspond to higher values of $Q^*(w)$. The straight line is the least squares regression line $y = 0.47x + 0.00197$.

as examples many well-studied risk and error measures in the literature, which are used as objective function in portfolio optimization, classification and regression. The special case of φ-divergence quadrangle provides the interpretation of the three problems as distributionally robust optimization. We visualize the risk identifier in portfolio optimization, classification and regression with a case study.

References

Ahmadi-Javid, A. (2011). Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155(3):1105–1123.

Ahmadi-Javid, A. (2012). Entropic value-at-risk: A new coherent risk measure. *Journal of Optimization Theory and Applications*, 155:1105–1123.

Ang, M., Sun, J., and Yao, Q. (2018). On the dual representation of coherent risk measures. *Annals of Operations Research*, 262:29–46.

Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical finance*, 9(3):203–228.

Bauschke, H. H. and Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, NY.

Bayraksan, G. and Love, D. K. (2015). Data-driven stochastic programming using phi-divergences. In *The operations research revolution*, pages 1–19. INFORMS.

Bellini, F., Klar, B., Müller, A., and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, 54:41–48.

Ben-Tal, A., Den Hertog, D., De Waegenaere, A., Melenberg, B., and Rennen, G. (2013). Robust solutions of optimization problems affected by uncertain probabilities. *Management Science*, 59(2):341–357.

Ben-Tal, A. and Teboulle, M. (1987). Penalty functions and duality in stochastic programming via φ-divergence functionals. *Mathematics of Operations Research*, 12(2):224–240.

Ben-Tal, A. and Teboulle, M. (2007). An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance*, 17(3):449–476.

Csiszár, I. (1963). Eine informations theoretische ungleichung und ihre anwendung auf den beweis der ergodizitat von markoffschen ketten. *Mayar Tud. Acad. Mat. Kutato Int. Kozl.*, 8:85–108.

Dacorogna, B. and Maréchal, P. (2008). The role of perspective functions in convexity, polyconvexity, rank-one convexity and separate convexity. *Journal of Convex Analysis*, 15(2):271–284.

Dommel, P. and Pichler, A. (2020). Convex Risk Measures based on Divergence. Papers 2003.07648, arXiv.org.

Föllmer, H. and Knispel, T. (2011). Entropic risk measures: Coherence vs. convexity, model ambiguity and robust large deviations. *Stochastics and Dynamics*, 11(02n03):333–351.

Föllmer, H. and Schied, A. (2011). *Stochastic finance: an introduction in discrete time*. Walter de Gruyter.

Fröhlich, C. and Williamson, R. C. (2022a). Risk measures and upper probabilities: Coherence and stratification. *arXiv preprint arXiv:2206.03183*.

Fröhlich, C. and Williamson, R. C. (2022b). Tailoring to the tails: Risk measures for fine-grained tail sensitivity. *arXiv preprint arXiv:2208.03066*.

Gotoh, J.-y. and Takeda, A. (2004). A linear classification model based on conditional geometric score.

Gotoh, J.-y. and Uryasev, S. (2017). Support vector machines based on convex risk functions and general norms. *Annals of Operations Research*, 249:301–328.

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

Krokhmal, P. A. (2007). Higher moment coherent risk measures.

Love, D. and Bayraksan, G. (2015). Phi-divergence constrained ambiguous stochastic programs for data-driven optimization. *Technical report, Department of Integrated Systems Engineering, The Ohio State University, Columbus, Ohio.*

Malandii, A., Kuzmenko, V., and Uryasev, S. (2024). Expectile risk quadrangles and applications. *Working Paper.*

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Markowitz, H. M. (1959). *Portfolio Selection: Efficient Diversification of Investments.* Yale University Press.

Morimoto, T. (1963). Markov processes and the h-theorem. *Journal of the Physical Society of Japan*, 18(3):328–331.

Pflug, G. C. and Ruszczynski, A. (2004). A risk measure for income processes.

Rockafellar, R. and Wets, R. J.-B. (1998). *Variational Analysis.* Springer Verlag, Heidelberg, Berlin, New York.

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

Rockafellar, R. T. (1976). Integral functionals, normal integrands and measurable selections. In *Nonlinear Operators and the Calculus of Variations*, pages 157–207. Springer Berlin Heidelberg, Berlin, Heidelberg.

Rockafellar, R. T. (1977). *Conjugate Duality and Optimization.* SIAM, Philadelphia, PA.

Rockafellar, R. T. (2020). Risk and utility in the duality framework of convex analysis. *From Analysis to Visualization: A Celebration of the Life and Legacy of Jonathan M. Borwein, Callaghan, Australia, September 2017*, pages 21–42.

Rockafellar, R. T. (2023). Distributional robustness, stochastic divergences, and the quadrangle of risk.

Rockafellar, R. T. and Royset, J. O. (2015). Measures of residual risk with connections to regression, risk tracking, surrogate models, and ambiguity. *SIAM Journal on Optimization*, 25(2):1179–1208.

Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of risk*, 2:21–42.

Rockafellar, R. T. and Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18(1-2):33–53.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, 10:51–74.

Rockafellar, R. T., Uryasev, S., and Zabarankin, M. (2008). Risk tuning with generalized linear regression. *Mathematics of Operations Research*, 33(3):712–729.

Royset, J. O. and Wets, R. J.-B. (2022). *An optimization primer*. Springer.

Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L. (2000). New support vector algorithms. *Neural computation*, 12(5):1207–1245.

Shapiro, A. (2017). Distributionally robust stochastic programming. *SIAM Journal on Optimization*, 27(4):2258–2275.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2014). *Lectures on Stochastic Programming: Modeling and Theory, Second Edition*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Sun, J., Yang, X., Yao, Q., and Zhang, M. (2020). Risk minimization, regret minimization and progressive hedging algorithms. *Mathematical Programming*, 181:509–530.

Takeda, A. and Sugiyama, M. (2008). ν-support vector machine as conditional value-at-risk minimization. In *Proceedings of the 25th international conference on Machine learning*, pages 1056–1063.

Zhang, T. and Zhou, Z.-H. (2014). Large margin distribution machine. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 313–322.

A Quadrangle Theorem

Rockafellar and Uryasev [2013] introduced measures of uncertainty that are built upon the concept of regularity, which is closely linked to convexity and closedness.

Uncertainty can be modeled via random variables and by studying and estimating the statistical properties of these random variables, we can estimate the risk in one form or the other. When the aim is to estimate the risk, it is convenient to think of the the random variable as 'loss' or 'cost'. There are various ways in which risk can be quantified and expressed. One such framework developed by Rockafellar and Uryasev [2013] is called the *Risk Quadrangle*, which is shown in Figure 4.

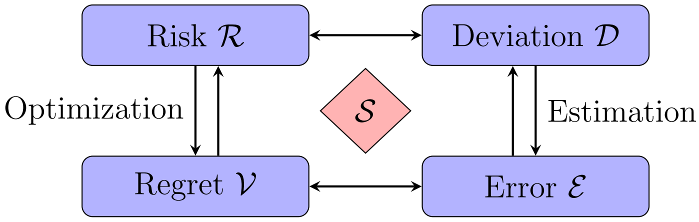

Figure 4: Risk Quadrangle Flowchart

The quadrangle begins from the upper left corner which depicts the *measure of risk* denoted by $\mathcal{R}$. It aggregates the uncertainty in losses into a numerical value $\mathcal{R}(X)$ by the inequality $\mathcal{R}(X) \leq C$ where C is the tolerance level for the risk. The next term is in the upper-right corner called the *measure of deviation* denoted by $\mathcal{D}$ and it quantifies the nonconstancy of the random variable. The lower-left corner depicts *measure of regret* denoted by $\mathcal{V}$. It stands for the net displeasure perceived in the potential mix of outcomes of a random variable "loss" which can be bad (> 0) or acceptable/good (≥ 0). The last measure is the *measure or error* which sits as the right-bottom of the quadrangle denoted by $\mathcal{E}$. Error quantifies the non-zeroness in the random variable.

Theorem A.1 (Quadrangle Theorem, Rockafellar and Uryasev [2013])**.** *The theorem states the following:*

(a) The centerness relations $\mathcal{D}(X) = \mathcal{R}(X) - \mathbb{E}[X]$ and $\mathcal{R}(X) = \mathbb{E}[X] + \mathcal{D}(X)$ give a one-to-one correspondence between regular measures of risk $\mathcal{R}$ and regular measures of deviation $\mathcal{D}$. In this correspondence, $\mathcal{R}$ is positively homogeneous if and only if $\mathcal{D}$ is positively homogeneous. On the other hand, $\mathcal{R}$ is monotonic if and only if $\mathcal{D}(X) \leq \sup X - \mathbb{E}[X]$ for all X.

(b) The relations $\mathcal{E}(X) = \mathcal{V}(X) - \mathbb{E}[X]$ and $\mathcal{V}(X) = \mathbb{E}[X] + \mathcal{E}(X)$ give a one-to-one correspondence between regular measures of regret V and regular measures of error $\mathcal{E}$. In this correspondence, $\mathcal{V}$ is positively homogeneous if and only if $\mathcal{E}$ is positively homogeneous. On the other hand, $\mathcal{V}$ is monotonic if and only if $\mathcal{E}(X) \leq |\mathbb{E}[X]|$ for $X \leq 0$.

(c) For any regular measure of regret $\mathcal{V}$, a regular measure of risk $\mathcal{E}$ is obtained by:

$$\mathcal{R}(X) = \min_{C \in \mathbb{R}} \{C + \mathcal{V}(X - C)\} \ .$$

If $\mathcal{V}$ is positively homogeneous, $\mathcal{R}$ is positively homogeneous. If $\mathcal{V}$ is monotonic, $\mathcal{R}$ is monotonic.

(d) *For any regular measure of error* $\mathcal{E}$, *a regular measure of deviation* $\mathcal{D}$ *is obtained by*

$$\mathcal{D}(X) = \min_{C \in \mathbb{R}} \{\mathcal{E}(X - C)\} \ .$$

If $\mathcal{E}$ *is positively homogeneous,* $\mathcal{D}$ *is positively homogeneous. If* $\mathcal{E}$ *satisfies the condition* $\mathcal{E}(X) \leq |\mathbb{E}[X]|$ *for* $X \leq 0$, *then* $\mathcal{D}$ *satisfies the condition* $\mathcal{D}(X) \leq \sup X - \mathbb{E}[X]$ *for all* X.

(e) *In both (c) and (d), as long as the expression being minimized is finite for some* C , *the set of* C *values for which the minimum is attained is a nonempty, closed, bounded interval. Moreover when* $\mathcal{V}$ *and* $\mathcal{E}$ *are paired as in (b), the interval comes out the same and gives the associated statistic:*

$$\operatorname*{argmin}_{C \in \mathbb{R}} \{C + \mathcal{V}(X - C)\} = \mathcal{S}(X) = \operatorname*{argmin}_{C \in \mathbb{R}} \{\mathcal{E}(X - C)\} \ .$$