Math. Program., Ser. B (2019) 174:525–552
https://doi.org/10.1007/s10107-018-1267-3

FULL LENGTH PAPER

Spectral risk measures: the risk quadrangle and optimal approximation

Drew P. Kouri[1]

Received: 21 April 2017 / Accepted: 22 March 2018 / Published online: 24 May 2018

Abstract We develop a general risk quadrangle that gives rise to a large class of spectral risk measures. The statistic of this new risk quadrangle is the average value-at-risk at a specific confidence level. As such, this risk quadrangle generates a continuum of error measures that can be used for superquantile regression. For risk-averse optimization, we introduce an optimal approximation of spectral risk measures using quadrature. We prove the consistency of this approximation and demonstrate our results through numerical examples.

Keywords Stochastic optimization · Risk measures · Regression · Quadrature · Average value-at-risk

Mathematics Subject Classification 49J20 · 49J55 · 49K20 · 49K45 · 90C15

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

This research was sponsored by DARPA EQUiPS Grant SNL 014150709.

✉ Drew P. Kouri
dpkouri@sandia.gov

[1] Optimization and Uncertainty Quantification, MS-1320, Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-1320, USA

1 Introduction

Countless science and engineering decisions are formulated as optimization problems, many of which are riddled with uncertainty. For such problems, it is critical that the optimal solutions not only account for, but are resilient to this uncertainty. Risk measures provide a convenient utility for quantifying the hazard associated with random objectives and constraints. In this paper, we consider the general risk-averse optimization problem

$$\min_{z\in Z_{\text{ad}}} \mathcal{R}(J(z)) \tag{1}$$

where $\mathcal{R}$ denotes a *spectral* risk measure [1], Z_{ad} is the feasible set of optimization variables and J is the random variable objective function. Problems of this form are frequently studied in financial application [12,13,19], but have recently gained much interest in engineering application, including PDE-constrained optimization [7,17,18,32]. Spectral risk measures[1] are law invariant, coherent [3] risk measures with the form

$$\mathcal{R}(X) = \int_0^1 \text{AVaR}_\beta(X)\,\mathrm{d}\mu(\beta) \tag{2}$$

where

$$\text{AVaR}_\alpha(X) := \frac{1}{1-\alpha}\int_\alpha^1 F_X^{-1}(\beta)\,\mathrm{d}\beta \tag{3}$$

is the average value-at-risk (AVaR)[2] with confidence level $\alpha \in [0,1)$ [29], $\text{AVaR}_1(X) = \text{ess sup}\ X$, F_X^{-1} is the quantile function of the random variable X and μ is a probability measure on the interval $[0,1]$. As shown in [31, Th. 6.43], under certain assumptions $\mathcal{R}$ defined in (2) is equivalent to $\mathcal{R}$ being a *comonotonic* risk measure.

The principle contributions of this paper are twofold: (i) we demonstrate that, under certain assumptions on μ, the associated spectral risk measure is generated by a risk quadrangle [30] whose statistic is $\text{AVaR}_\alpha(X)$ with an appropriate confidence level α; and (ii) we develop a provably optimal quadrature approximation for spectral risk measures whose defining measures μ are absolutely continuous with respect to the Lebesgue measure. The risk quadrangle provides a rigorous mathematical connection between stochastic optimization and statistical estimation, through the definition of generalized regret, risk, deviation and error measures. Using the associated error measures for our spectral risk quadrangle, one can perform *superquantile regression* [28], permitting the injection of risk-aversion into regression models. See [20] for a related approach for risk-averse regression modeling. In fact, the risk quadrangle developed here provides a continuum of possible error measures, and thus regression models, that approximate conditional versions of AVaR.

This paper is structured as follows. First, we introduce notation and review law-invariant coherent risk measures. Then, we develop the appropriate assumptions on the

[1] The spectral risk measures form a subset of the more general class of distortion risk measures [9]. In addition, the authors in [27] refer to spectral risk measures as mixed superquantile risk measures.

[2] Also called conditional tail expectation, conditional value-at-risk, expected shortfall, expected tail loss and superquantile.

spectral risk measures that give rise to risk quadrangles. Following our risk quadrangle results, we discuss the numerical evaluation of spectral risk measures. In particular, we introduce a provably optimal quadrature approximation. Using this quadrature approximation, we prove consistency of minimizers and stationary points of (1) when $\mathcal{R}$ is replaced by our quadrature approximation. We conclude with numerical examples that confirm our results.

2 Notation and terminology

For any Banach space V, we denote the norm on V by $\|\cdot\|_V$ and the topological dual space of V by V^*. That is, V^* consists of all continuous linear functionals acting on V. For any $v \in V$ and $v^* \in V^*$, we denote the action of v^* on v by $\langle v^*, v\rangle_{V^*,V}$. Moreover, for any sequence $\{v_n\} \subset V$, we say that v_n strongly converges to v if

$$\lim_{n\to\infty} \|v_n - v\|_V = 0,$$

i.e., v_n converges to v with respect to the norm topology on V. We denote strong convergence by $v_n \to v$. We say that v_n weakly converges to $v \in V$ if

$$\lim_{n\to\infty} \langle v^*, v_n - v\rangle_{V^*,V} = 0 \quad \forall\, v^* \in V^*,$$

i.e., v_n converges to v with respect to the weak topology on V. We denote weak convergence by $v_n \rightharpoonup v$. In addition, for any sequence $\{v_n^*\} \subset V^*$, we say that v_n^* weakly* converges to $v^* \in V^*$ if

$$\lim_{n\to\infty} \langle v_n^* - v^*, v\rangle_{V^*,V} = 0 \quad \forall\, v \in V,$$

i.e., v_n^* converges to v^* with respect to the weak* topology on V^*. We denote weak* convergence by $v_n^* \rightharpoonup^* v^*$. Note that if V is reflexive, then weak* convergence on V^* is equivalent to weak convergence on V^*. Finally, for any subset $A \subset V$, we say that A is strongly (weakly) closed if it contains all of its strong (weak) accumulation points. If A is convex, then A is weakly closed if and only if A is strongly closed. Analogously, a set $B \subset V^*$ is weakly* closed if it contains all of its weak* accumulation points. See [5, Ch. 2] for more details.

3 Risk measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a nonatomic probability space where Ω is the set of possible outcomes, $\mathcal{F} \subseteq 2^\Omega$ is a σ-algebra of events, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a nonatomic probability measure. Moreover, let $\mathcal{X} := L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty)$ denote the space of real-valued $\mathcal{F}$-measurable functions with p finite moments, i.e., $\int_\Omega |X|^p(\omega)\mathrm{d}\mathbb{P}(\omega) < \infty$. Throughout, we denote the expected value of a random variable $X \in \mathcal{X}$ by $\mathbb{E}[X] := \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}(\omega)$ and the cumulative distribution function (cdf) of X as $F_X(x) := \mathbb{P}(X \le x)$. Moreover, we denote the Lebesgue measure on $[0, 1]$ by m and set $\mathrm{d}m(\beta) = \mathrm{d}\beta$ to simplify notation.

A functional $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$ is a *coherent risk measure* if it satisfies the following four axioms [3]:

(R1) *Monotonicity:* If $X, X' \in \mathcal{X}$ satisfy $X \geq X'$ $\mathbb{P}$-a.e., then $\mathcal{R}(X) \geq \mathcal{R}(X')$;
(R2) *Subadditivity:* If $X, X' \in \mathcal{X}$, then $\mathcal{R}(X + X') \leq \mathcal{R}(X) + \mathcal{R}(X')$;
(R3) *Translation Equivariance:* If $t \in \mathbb{R}$ and $X \in \mathcal{X}$, then $\mathcal{R}(X + t) = \mathcal{R}(X) + t$;
(R4) *Positive Homogeneity:* If $t \geq 0$ and $X \in \mathcal{X}$, then $\mathcal{R}(tX) = t\mathcal{R}(X)$.

In addition to the definition of coherent risk measures, a functional $\mathcal{R} : \mathcal{X} \to \overline{\mathbb{R}}$ is *law invariant* if $X,\ X' \in \mathcal{X}$ satisfying

$$F_X(x) = F_{X'}(x) \quad \forall\, x \in \mathbb{R} \quad \Longrightarrow \quad \mathcal{R}(X) = \mathcal{R}(X'). \tag{4}$$

A popular law-invariant coherent risk measure is the average value-at-risk (AVaR), defined in (3). If the cdf, F_X, of the random variable $X \in \mathcal{X}$ is continuous, then $\text{AVaR}_\alpha(X)$ is the expectation of X conditioned on the event that X is larger than its α-quantile, $F_X^{-1}(\alpha) := \inf\{\eta : \mathbb{P}(X \leq \eta) \geq \alpha\}$. In general, the α-AVaR of X can be written in the equivalent form

$$\text{AVaR}_\alpha(X) = \inf_{t \in \mathbb{R}} \left\{ t + \frac{1}{1-\alpha}\mathbb{E}[(X-t)_+] \right\} \tag{5}$$

for $\alpha \in [0, 1)$ where $(\cdot)_+ := \max\{0, \cdot\}$. For what follows, it is important to recall that the mapping $\alpha \mapsto \text{AVaR}_\alpha(X) : [0, 1] \to \overline{\mathbb{R}}$ is continuous and increasing on $[0, 1)$ with

$$\text{AVaR}_0(X) = \mathbb{E}[X] \quad \text{and} \quad \text{AVaR}_1(X) = \lim_{\alpha \to 1} \text{AVaR}_\alpha(X) = \text{ess sup}\, X$$

where $\text{AVaR}_1(X)$ is possibly infinite. In fact, $\alpha \mapsto \text{AVaR}_\alpha(X)$ is strictly increasing on the interval $[0, 1 - \mathbb{P}(X = \text{ess sup } X))$ and constant on the interval $[1 - \mathbb{P}(X = \text{ess sup } X), 1]$ for nondegenerate[3] $X \in \mathcal{X}$ [27].

In this paper, we are interested in the specific class of law-invariant coherent risk measures with the form (2) where μ is a fixed probability measure on $[0, 1]$. Many popular risk measures fit into this class. For example, if $\mu = \delta_\alpha$ where δ_α denotes the point mass centered at α for fixed $0 \leq \alpha \leq 1$, then

$$\mathcal{R}(X) = \text{AVaR}_\alpha(X).$$

This can be generalized to a convex combination of point masses $\mu = \lambda_1\delta_{\alpha_1} + \cdots + \lambda_n\delta_{\alpha_n}$ where $\lambda_i > 0$ for all $i = 1, \ldots, n$ with $\lambda_1 + \cdots + \lambda_n = 1$ and $0 \leq \alpha_1 < \cdots < \alpha_n \leq 1$. In this case, the corresponding risk measure is the mixture of AVaRs

$$\mathcal{R}(X) = \lambda_1 \text{AVaR}_{\alpha_1}(X) + \cdots + \lambda_n \text{AVaR}_{\alpha_n}(X).$$

[3] A random variable $X \in \mathcal{X}$ is degenerate if $X = \mathbb{E}[X]$ $\mathbb{P}$-a.e.

	Risk	$\mathcal{R}$	$\longleftrightarrow$	$\mathcal{D}$	Deviation	
Optimization		$\uparrow\downarrow$	$\mathcal{S}$	$\downarrow\uparrow$		Estimation
	Regret	$\mathcal{V}$	$\longleftrightarrow$	$\mathcal{E}$	Error	

$$\mathcal{R}(X) = \mathbb{E}[X] + \mathcal{D}(X) \qquad \mathcal{D}(X) = \mathcal{R}(X) - \mathbb{E}[X]$$
$$\mathcal{V}(X) = \mathbb{E}[X] + \mathcal{E}(X) \qquad \mathcal{E}(X) = \mathcal{V}(X) - \mathbb{E}[X]$$
$$\mathcal{R}(X) = \inf_t \{t + \mathcal{V}(X - t)\} \qquad \mathcal{D}(X) = \inf_t \mathcal{E}(X - t)$$
$$\mathcal{S}(X) = \arg\min_t \{t + \mathcal{V}(X - t)\} = \arg\min_t \mathcal{E}(X - t)$$

Fig. 1 The risk quadrangle provides rigorous mathematical connections between stochastic optimization and statistical estimation through the definitions of measures of risk, deviation, regret and error

One final example is the risk measure that results from the superquantile risk quadrangle [24,25]. To generate this risk measure, we define $\mu = (1-\alpha)^{-1} m(\cdot \cap [\alpha, 1])$ where $0 \leq \alpha < 1$ which produces

$$\mathcal{R}(X) = \frac{1}{1-\alpha} \int_\alpha^1 \mathrm{AVaR}_\beta(X)\, \mathrm{d}\beta.$$

We note that any spectral risk measure (2), with defining measure μ, can be equivalently evaluated as

$$\mathcal{R}(X) = \int_0^1 h(\beta) F_X^{-1}(\beta)\, \mathrm{d}\beta \tag{6}$$

where the *spectral function* h (i.e., $h : [0, 1) \to [0, +\infty)$ is right continuous, monotonically increasing and satisfies $\int_0^1 h(\beta)\, \mathrm{d}\beta = 1$) is defined as

$$h(\beta) := \int_0^\beta \frac{1}{1-\alpha}\, \mathrm{d}\mu(\alpha).$$

See [31, Ch. 6.3.4-5] for more details. Equation (6) could form the basis for numerical approximation of $\mathcal{R}(X)$ using, e.g., quadrature. However, in order to evaluate $\mathcal{R}(X)$, one must be able to accurately evaluate $F_X^{-1}(\beta)$ at some point β. Moreover, the resulting approximate optimization problem need not be convex even if (1) is (i.e., the approximate objective function is a linear combination of quantiles). Instead, we utilize (2) and (5) to form the basis of our numerical approximation. By approximating (2) using an appropriate quadrature rule, we arrive at a mixture of AVaRs. We can then incorporate t_i in (5) corresponding to each confidence level (quadrature abscissa) α_i as an additional variable in our optimization problem (1). In this way, we are able to avoid directly computing quantiles. This approach additionally results in a convex optimization problem when (1) is convex.

4 The risk quadrangle

The risk quadrangle, introduced in [30], provides a rigorous connection between statistical estimation and risk-averse optimization by furnishing fundamental relationships

connecting regular measures of risk, deviation, error and regret (see Fig. 1). The risk quadrangle was originally analyzed for $\mathcal{X} = L^2(\Omega, \mathcal{F}, \mathbb{P})$. However, the definitions of regular measures of risk, deviation, error and regret, and the fundamental relations between them hold for $\mathcal{X} = L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty)$.[4] We develop a regular measure of regret, defined on $\mathcal{X} = L^p(\Omega, \mathcal{F}, \mathbb{P})$ for $p \in [1, \infty)$, that gives rise to a class of spectral risk measures. We then apply the fundamental risk quadrangle relationships (Fig. 1) to obtain risk, error and deviation measures. A regular measure of regret (as defined in [30]), $\mathcal{V} : \mathcal{X} \to \overline{\mathbb{R}}$, is proper, closed, convex and satisfies $\mathcal{V}(0) = 0$, $\mathcal{V}(X) > \mathbb{E}[X]$ for all $0 \not\equiv X \in \mathcal{X}$ and

$$\lim_{k\to\infty} \{\mathcal{V}(X_k) - \mathbb{E}[X_k]\} = 0 \quad \Longrightarrow \quad \lim_{k\to\infty} \mathbb{E}[X_k] = 0. \tag{7}$$

We mention here that the authors of [26] have extend the Quadrangle Theorem of [30] to account for measures of regret that do not satisfy condition (7).

The authors in [24] construct the superquantile quadrangle using the regret measure

$$\mathcal{V}(X) = \frac{1}{1-\alpha} \int_0^1 (\mathrm{AVaR}_\beta(X))_+ \, \mathrm{d}\beta. \tag{8}$$

We use this regret measure to motivate our generalization. Let $\mathcal{B} \subseteq 2^{[0,1]}$ be the Borel σ-algebra on $[0, 1]$ and $\nu : \mathcal{B} \to [0, \infty)$ be a finite positive measure on $[0, 1]$ satisfying the following assumptions.

Assumption 1 $\nu : \mathcal{B} \to [0, \infty)$ is a positive and finite measure satisfying:

1. $\nu([0, 1]) = c \in (1, \infty)$;
2. $\nu((0, 1]) > 0$;
3. $\nu(\{1\}) = 0$ and $\int_0^1 (1-\beta)^{-\frac{1}{p}} \, \mathrm{d}\nu(\beta) < \infty$.

Our generalization of the superquantile regret is

$$\mathcal{V}_\nu(X) = \int_0^1 (\mathrm{AVaR}_\beta(X))_+ \, \mathrm{d}\nu(\beta). \tag{9}$$

Note that, if $\nu = (1-\alpha)^{-1} m$, then (9) is exactly the superquantile regret. We observe that $\mathcal{V}_\nu$ defined in (9) is convex, positive homogeneous and monotonic. Moreover, Assumption 1.1-2 ensures that $\mathcal{V}_\nu$ satisfies

$$\mathcal{V}_\nu(0) = 0 \quad \text{and} \quad \mathcal{V}_\nu(X) > \mathbb{E}[X] \quad \forall\, X \in \mathcal{X},\ X \not\equiv 0, \tag{10}$$

since $\mathrm{AVaR}_\beta(X) > \mathbb{E}[X]$ with $\beta \in (0, 1]$ for all nondegenerate $X \in \mathcal{X}$ and $\mathcal{V}_\nu(X) = c(\mathbb{E}[X])_+ > \mathbb{E}[X]$ for all degenerate $X \in \mathcal{X}$ with $\mathbb{E}[X] \neq 0$. Since $\mathcal{V}_\nu$ is proper, convex and monotonic, Proposition 2.5 in [10] ensures $\mathcal{V}_\nu$ is norm continuous if it is

[4] The results follow since a convex set in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ with $p \in [1, \infty)$ is weakly closed if and only if it is strongly closed.

finite. In the subsequent proposition, we show that $\mathcal{V}_\nu$ is finite. A similar result was proved in [27, Prop. 2.3] for $p = 2$.

Proposition 1 *Let Assumption 1 hold. Then $\mathcal{V}_\nu$ defined in (9) is finite and hence (norm) continuous on $\mathcal{X}$.*

Proof The proof of this result is similar to that of Proposition 1 in [28]. Let $\mathbb{1}_\alpha$ denote the characteristic function of the interval $[F_X^{-1}(\alpha), \infty)$ for a random variable $X \in \mathcal{X}$. As in the proof of [28, Prop. 1], if F_X^{-1} is continuous at α, then we have that

$$(1-\alpha)\mathrm{AVaR}_\alpha(X - \mathbb{E}[X]) = \mathbb{E}[(X - \mathbb{E}[X])\mathbb{1}_\alpha] \le \mathbb{E}[|X - \mathbb{E}[X]|^p]^{\frac{1}{p}} \mathbb{E}[\mathbb{1}_\alpha^q]^{\frac{1}{q}}. \quad (11)$$

where $\frac{1}{p} + \frac{1}{q} = 1$. Here, we have used Hölder's inequality in place of Cauchy-Schwartz. Rearranging (11) gives

$$\mathrm{AVaR}_\alpha(X) \le \mathbb{E}[X] + \frac{1}{(1-\alpha)^{\frac{1}{p}}} \sigma_p(X) =: \theta_X(\alpha) \quad (12)$$

where $\sigma_p(X) := \mathbb{E}[|X - \mathbb{E}[X]|^p]^{\frac{1}{p}} < \infty$ for $X \in \mathcal{X}$. Since F_X^{-1} is continuous at all but countably many α, the continuity of both sides of (12) ensures the inequality holds for all $\alpha \in (0, 1)$.

Now, since $\alpha \mapsto \mathrm{AVaR}_\alpha(X)$ is continuous, it is $\mathcal{B}$-measurable. Thus, the integral on the right-hand side of (9) is well defined and (12) yields the bound

$$\mathcal{V}_\nu(X) \le \int_0^1 (\theta_X(\beta))_+ \, \mathrm{d}\nu(\beta).$$

If X is degenerate, then $\mathcal{V}_\nu(X) = c(\mathbb{E}[X])_+$ is finite. On the other hand, if X is not degenerate, then $\sigma_p(X) > 0$. In this case, let $\bar{\alpha} = \inf\{\alpha \in [0, 1) : \theta_X(\alpha) > 0\}$, then $\theta_X(\alpha) \le 0$ for all $\alpha \in [0, \bar{\alpha})$ and $(\theta_X(\alpha))_+ = \theta_X(\alpha)$ is increasing for all $\alpha \in [\bar{\alpha}, 1)$. Thus,

$$\int_0^1 (\theta_X(\beta))_+ \, \mathrm{d}\nu(\beta) = \nu([\bar{\alpha}, 1])\mathbb{E}[X] + \sigma_p(X) \int_{\bar{\alpha}}^1 \frac{1}{(1-\beta)^{\frac{1}{p}}} \, \mathrm{d}\nu(\beta).$$

Hence, $\mathcal{V}_\nu$ is finite by Assumption 1.3 and continuous by [10, Prop. 2.5]. □

Proposition 1 ensures that $\mathcal{V}_\nu$ is continuous and, hence, closed. Therefore, under appropriate assumptions on ν, the aforementioned properties guarantee that $\mathcal{V}_\nu$ satisfies all axioms of a regular measure of regret with the exception of (7). We will use the next proposition to prove (7).

Proposition 2 *Suppose $\nu : \mathcal{B} \to [0, \infty)$ satisfies Assumption 1. Then,*

$$\mathcal{V}_\nu(X) - \mathbb{E}[X] \ge \min\{1, c - 1\}|\mathbb{E}[X]| \quad \forall\, X \in \mathcal{X}.$$

Proof We consider two cases: (i) $\mathbb{E}[X] \geq 0$ and (ii) $\mathbb{E}[X] < 0$. Since $\mathrm{AVaR}_\alpha(X) \geq \mathbb{E}[X]$ for all α, in case (i) we have that

$$\int_0^1 (\mathrm{AVaR}_\beta(X))_+ \, \mathrm{d}\nu(\beta) - \mathbb{E}[X] \geq (\nu([0,1]) - 1)\mathbb{E}[X] = (c-1)|\mathbb{E}[X]|$$

and in case (ii) we have that

$$\int_0^1 (\mathrm{AVaR}_\beta(X))_+ \, \mathrm{d}\nu(\beta) - \mathbb{E}[X] \geq -\mathbb{E}[X] = |\mathbb{E}[X]|.$$

This proves the desired result. □

Remark 1 Assumption 1 ensures that $\min\{1, c-1\} > 0$ and hence

$$\mathcal{V}_\nu(X_k) - \mathbb{E}[X_k] \geq \min\{1, c-1\}|\mathbb{E}[X_k]| \geq 0$$

for any sequence $\{X_k\} \subset \mathcal{X}$. Therefore,

$$\lim_{k\to\infty} \{\mathcal{V}_\nu(X_k) - \mathbb{E}[X_k]\} = 0 \implies \lim_{k\to\infty} |\mathbb{E}[X_k]| = 0 \implies (7) \text{ holds}$$

and $\mathcal{V}_\nu$ is a regular measure of regret.

Remark 2 In the case that the measure $\nu([0,1]) = c \leq 1$, we can augment ν as $\nu_1 = \delta_0 + \nu$. This new measure satisfies $\nu_1([0,1]) = 1 + c$ and thus Proposition 2 ensures that $\mathcal{V}_{\nu_1}$ is a regular measure of regret. Another approach would be to set $\nu_2 = \bar{c}\nu$ with $\bar{c} > c^{-1}$. Again, Proposition 2 ensures that $\mathcal{V}_{\nu_2}$ is a regular measure of regret. For example, fix $0 < \alpha < 1$ and let $\nu = \delta_\alpha$. Then, the regret measures corresponding to ν_1 and ν_2 are

$$\mathcal{V}_{\nu_1}(X) = (\mathbb{E}[X])_+ + \int_0^1 (\mathrm{AVaR}_\beta(X))_+ \, \mathrm{d}\nu(\beta)$$
$$\mathcal{V}_{\nu_2}(X) = \bar{c} \int_0^1 (\mathrm{AVaR}_\beta(X))_+ \, \mathrm{d}\nu(\beta),$$

respectively. However, these two approaches give rise to different risk measures.

Under Assumption 1, Propositions 1 and 2 ensure that $\mathcal{V}_\nu$ is a regular measure of regret. Applying the risk quadrangle relationships (see Fig. 1) yields the error measure

$$\mathcal{E}_\nu(X) = \mathcal{V}_\nu(X) - \mathbb{E}[X]. \tag{13}$$

Clearly, $\mathcal{E}_\nu$ is closed, convex and satisfies $\mathcal{E}_\nu(0) = 0$, $\mathcal{E}_\nu(X) > 0$ for all $0 \not\equiv X \in \mathcal{X}$ and (7) ensures

$$\lim_{k\to\infty} \mathcal{E}_\nu(X_k) = 0 \implies \lim_{k\to\infty} \mathbb{E}[X_k] = 0.$$

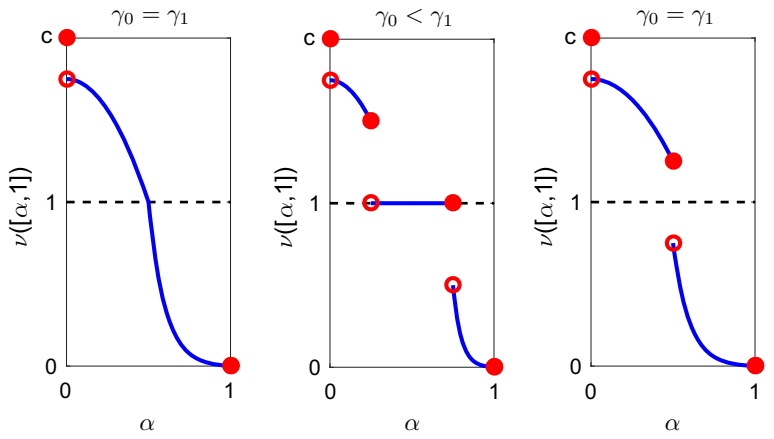

Fig. 2 Examples of the left continuous function $\alpha \mapsto \nu([\alpha, 1])$

Hence, $\mathcal{E}_\nu$ is a regular measure of error [30, pg. 43]. Additionally, the corresponding measures of risk and deviation are

$$\mathcal{R}_\nu(X) = \inf_{t\in\mathbb{R}}\{t + \mathcal{V}_\nu(X - t)\} \quad \text{and} \quad \mathcal{D}_\nu(X) = \inf_{t\in\mathbb{R}} \mathcal{E}_\nu(X - t).$$

Finally, the associated statistic is

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

We will now determine an explicit form for the statistic $\mathcal{S}_\nu$ which we will then use to derive an explicit representation of the risk $\mathcal{R}_\nu$ and the deviation $\mathcal{D}_\nu$. The statistic depends on the function $\alpha \mapsto \nu([\alpha, 1])$ (see Fig. 2). In the context of regression, the statistic is the (conditional) quantity to be approximated.

Theorem 2 *Let Assumption 1 hold and define the scalars*

$$\gamma_0 = \sup\{\, t \in [0, 1] \,:\, \nu([t, 1]) > 1 \,\} \quad \text{and}$$
$$\gamma_1 = \inf\{\, t \in [0, 1] \,:\, \nu([t, 1]) < 1 \,\}.$$

Then $\mathcal{S}_\nu(X) = [AVaR_{\gamma_0}(X), AVaR_{\gamma_1}(X)]$ *for all* $X \in \mathcal{X}$. *In particular, if* $\gamma_0 = \gamma_1$, *then* $\mathcal{S}_\nu(X) = AVaR_{\gamma_0}(X)$.

Proof Define $\phi(t) = (t + \mathcal{V}_\nu(X - t))$, then ϕ is convex and continuous since $\mathcal{V}_\nu$ is. First note that for all $t \le \mathbb{E}[X]$, $\phi(t)$ is strictly decreasing and for all $t \ge \operatorname{ess\,sup} X$, $\phi(t)$ is strictly increasing. Thus, it suffices to consider $t = \mathrm{AVaR}_\gamma(X)$ for $0 < \gamma < 1$, in which case we have

$$\phi(\mathrm{AVaR}_\gamma(X)) = \mathrm{AVaR}_\gamma(X) + \int_0^1 (\mathrm{AVaR}_\beta(X - \mathrm{AVaR}_\gamma(X)))_+ \,\mathrm{d}\nu(\beta)$$
$$= (1 - \nu([\gamma, 1]))\mathrm{AVaR}_\gamma(X) + \int_\gamma^1 \mathrm{AVaR}_\beta(X)\,\mathrm{d}\nu(\beta). \tag{14}$$

For the subsequent results, it will be convenient to decompose ν. For any $\gamma \in [0, 1)$, we have the decomposition $\nu = \nu_\gamma + c_\gamma \delta_\gamma$ where $\nu_\gamma(\{\gamma\}) = 0$ and $c_\gamma := \nu(\{\gamma\}) \geq 0$. Now, note that $\mathrm{AVaR}_\gamma(X) = \operatorname{ess\,sup} X$ for all $\gamma \in [1 - \mathbb{P}(X = \operatorname{ess\,sup} X), 1]$. Therefore, we can exclude this case. We first choose any $\gamma^+ \in [\gamma_1, 1 - \mathbb{P}(X = \operatorname{ess\,sup} X))$. Since $\gamma^+ < 1 - \mathbb{P}(X = \operatorname{ess\,sup} X)$, we have that $\mathrm{AVaR}_{\gamma^+}(X) < \mathrm{AVaR}_\gamma(X)$ for any $\gamma \in (\gamma^+, 1 - \mathbb{P}(X = \operatorname{ess\,sup} X))$ (i.e., $\alpha \mapsto \mathrm{AVaR}_\alpha(X)$ is strictly increasing in this regime). Then by (14) and the decomposition $\nu = \nu_\gamma + c_\gamma \delta_\gamma$, we have

$$\phi(\mathrm{AVaR}_\gamma(X)) = (1 - \nu([\gamma^+, 1]))\mathrm{AVaR}_\gamma(X) + \int_{\gamma^+}^1 \mathrm{AVaR}_\beta(X)\,\mathrm{d}\nu(\beta)$$
$$+ \int_{\gamma^+}^{\gamma} (\mathrm{AVaR}_\gamma(X) - \mathrm{AVaR}_\beta(X))\,\mathrm{d}\nu_\gamma(\beta) > \phi(\mathrm{AVaR}_{\gamma^+}(X)).$$

Therefore, ϕ is strictly increasing on $[\mathrm{AVaR}_{\gamma_1}(X), \mathrm{AVaR}_1(X))$. In a similar fashion, suppose $\gamma^- \in (0, \gamma_0]$ and take any $\gamma \in [0, \gamma^-)$, then $\mathrm{AVaR}_\gamma(X) < \mathrm{AVaR}_{\gamma^-}(X)$ (again, $\alpha \mapsto \mathrm{AVaR}_\alpha(X)$ is strictly increasing in this regime). Then by (14) and the decomposition $\nu = \nu_{\gamma^-} + c_{\gamma^-}\delta_{\gamma^-}$, we have

$$\phi(\mathrm{AVaR}_\gamma(X)) = (1 - \nu([\gamma^-, 1]))\mathrm{AVaR}_\gamma(X) + \int_{\gamma^-}^1 \mathrm{AVaR}_\beta(X)\,\mathrm{d}\nu(\beta)$$
$$+ \int_{\gamma}^{\gamma^-} (\mathrm{AVaR}_\beta(X) - \mathrm{AVaR}_\gamma(X))\,\mathrm{d}\nu_{\gamma^-}(\beta) > \phi(\mathrm{AVaR}_{\gamma^-}(X)).$$

Therefore, ϕ is strictly decreasing on $[\mathrm{AVaR}_0(X), \mathrm{AVaR}_{\gamma_0}(X)]$. Hence, it suffices to understand the behavior of ϕ on $[\mathrm{AVaR}_{\gamma_0}(X), \mathrm{AVaR}_{\gamma_1}(X)]$. If $\gamma_0 < \gamma_1$, then for any $\gamma \in (\gamma_0, \gamma_1)$ we have that $\nu([\gamma, 1]) = 1$ and $\nu([\gamma, \gamma_1)]) = 0$. Thus,

$$\phi(\mathrm{AVaR}_\gamma(X)) = 0 + \int_\gamma^1 \mathrm{AVaR}_\beta(X)\,\mathrm{d}\nu(\beta) = \int_{\gamma_1}^1 \mathrm{AVaR}_\beta(X)\,\mathrm{d}\nu(\beta).$$

By continuity of ϕ and $\alpha \mapsto \mathrm{AVaR}_\alpha(X)$, we have that ϕ is constant on $\mathcal{S}_\nu(X) = [\mathrm{AVaR}_{\gamma_0}(X), \mathrm{AVaR}_{\gamma_1}(X)]$ and the desired result follows. To conclude, when $\gamma_0 = \gamma_1$, the interval $\mathcal{S}_\nu(X)$ becomes the singleton $\{\mathrm{AVaR}_{\gamma_0}(X)\}$. □

As Theorems 2 indicates, the statistic $\mathcal{S}_\nu$ is the average value-at-risk at a confidence level γ determined by the defining measure ν. Therefore, $\mathcal{E}_\nu$ provides a family of error measures capable of performing superquantile regression. It is unclear whether or not this increased generality provides any benefit over the error measure developed in

[28], for example, with tracking AVaR in the sense of [28, Sect. 3.2]. On the other hand, (14) provides the explicit form of $\mathcal{R}_\nu$. Namely,

$$\mathcal{R}_\nu(X) = (1 - \nu([\gamma, 1]))\text{AVaR}_\gamma(X) + \int_\gamma^1 \text{AVaR}_\beta(X)\, \mathrm{d}\nu(\beta) \tag{15}$$

for any $\gamma \in [\gamma_0, \gamma_1]$. By setting $\gamma = \gamma_0$, we have that

$$\mathcal{R}_\nu(X) = (1 - \nu_{\gamma_0}([\gamma_0, 1]))\text{AVaR}_{\gamma_0}(X) + \int_{\gamma_0}^1 \text{AVaR}_\beta(X)\, \mathrm{d}\nu_{\gamma_0}(\beta)$$

where $\nu = \nu_{\gamma_0} + c_{\gamma_0}\delta_{\gamma_0}$ with $\nu_{\gamma_0}(\{\gamma_0\}) = 0$ and $c_{\gamma_0} = \nu(\{\gamma_0\}) \geq 0$. Furthermore, since AVaR is translation equivariant, the deviation measure is

$$\mathcal{D}_\nu(X) = (1 - \nu_{\gamma_0}([\gamma_0, 1]))\text{AVaR}_{\gamma_0}(X - \mathbb{E}[X]) + \int_{\gamma_0}^1 \text{AVaR}_\beta(X - \mathbb{E}[X])\, \mathrm{d}\nu_{\gamma_0}(\beta).$$

Clearly, since $\nu([\gamma_0, 1]) \leq 1$, $\mathcal{R}_\nu$ in (15) is a spectral risk measure with defining measure

$$\mu = \nu_{\gamma_0}(\cdot \cap [\gamma_0, 1]) + (1 - \nu_{\gamma_0}([\gamma_0, 1]))\delta_{\gamma_0}. \tag{16}$$

In addition, any spectral risk measure whose defining probability measure, μ, satisfies the appropriate integrability requirements in Assumption 1.3 can be decomposed as in (16). That is, let $\eta = \sup\{\, t \in [0, 1] \,:\, \mu([0, t)) = 0\,\}$. Then μ has the form (16) with $\gamma_0 = \eta$. In this case, the associated spectral risk measure is generated through the risk quadrangle defined by the regret measure (9) with measure, for example, $\nu = \mu + c_\eta\delta_\eta$ for any $c_\eta > 0$. The statistic corresponding to this quadrangle is $\mathcal{S}_\nu(X) = \text{AVaR}_\eta(X)$.

5 Computation

In this section, we study the evaluation of a spectral risk measure $\mathcal{R}$ at X (or an approximation of X). Similar to the results in Section 5 of [28], we can derive an explicit form for $\mathcal{R}$ applied to a discrete approximation of X. We briefly discuss this derivation and the associated challenges that arise when applied to optimization problem (1). In the remainder of this section, we consider arbitrary $X \in \mathcal{X}$ (not necessarily discretely distributed). We introduce and analyze a quadrature approximation of $\mathcal{R}$ and investigate the affects of this approximation on the solutions of our target optimization problem. Throughout this section, we assume the measure μ in the definition of $\mathcal{R}$ is absolutely continuous with respect to the Lebesgue measure m and satisfies Assumption 1.3. Under these assumptions, μ admits a Lebesgue density function w satisfying $\mathrm{d}\mu(\beta) = w\, \mathrm{d}\beta$ and $\mathcal{R}(X)$ is finite for all $X \in \mathcal{X}$.

5.1 Approximation via discrete random variables

Let $\widehat{X}$ be a discretely distributed random variable that approximates X. We assume that $\widehat{X}$ has atoms $\{x_1, \ldots, x_N\}$ with associated probabilities $p_i = \mathbb{P}(\widehat{X} = x_i)$ for $i = 1, \ldots, N$, $N \in \mathbb{N}$. Throughout this subsection, we assume that σ is a permutation of the indices $\{1, \ldots, N\}$ such that $x_{\sigma(1)} < \cdots < x_{\sigma(N)}$. The cdf, $F_{\widehat{X}}$, is piecewise constant with range $\{0 = \beta_0, \beta_1, \ldots, \beta_N = 1\}$ where $\beta_i = p_{\sigma(1)} + \cdots + p_{\sigma(i)}$ for $i = 1, \ldots, N$. With these definitions, we can write

$$\mathcal{R}(\widehat{X}) = \sum_{i=1}^{N} \int_{\beta_{i-1}}^{\beta_i} \mathrm{AVaR}_\beta(\widehat{X}) \, \mathrm{d}\mu(\beta).$$

For any $\beta \in (\beta_{i-1}, \beta_i)$, the quantile $F^{-1}_{\widehat{X}}(\beta)$ is constant. Therefore, the integral over $(\beta_{N-1}, 1)$ is equal to $M = \mu([\beta_{N-1}, 1])x_{\sigma(N)}$. Additionally, if we define $t_i = F^{-1}_{\widehat{X}}(\beta) = x_{\sigma(i)}$ with $\beta \in (\beta_{i-1}, \beta_i)$, then

$$\begin{aligned}
\mathcal{R}(\widehat{X}) &= \sum_{i=1}^{N-1} \int_{\beta_{i-1}}^{\beta_i} \left\{ t_i + \frac{1}{1-\beta} \mathbb{E}[(\widehat{X} - t_i)_+] \right\} \mathrm{d}\mu(\beta) + M \\
&= \sum_{i=1}^{N} \mu([\beta_{i-1}, \beta_i]) \, t_i + \mathbb{E}[(\widehat{X} - t_i)_+] \int_{\beta_{i-1}}^{\beta_i} \frac{1}{1-\beta} \, \mathrm{d}\mu(\beta) + M.
\end{aligned}$$

Now, evaluating $A_i := \mathbb{E}[(\widehat{X} - t_i)_+] = \sum_{j=1}^{N} p_j (x_j - t_i)_+$ for $i = 1, \ldots, N-1$ gives

$$\mathcal{R}(\widehat{X}) = \sum_{i=1}^{N-1} \left\{ \mu([\beta_{i-1}, \beta_i]) \, t_i + A_i \int_{\beta_{i-1}}^{\beta_i} \frac{1}{1-\beta} \, \mathrm{d}\mu(\beta) \right\} + M.$$

Although t_i can be incorporated as optimization variables, this formulation presents a challenge when solving the optimization problem (1) since β_i, A_i and M depend on a permutation of the scenarios of $X = J(z)$ and hence on the optimization variables $z \in Z_{\mathrm{ad}}$. The next approach circumvents this issue.

5.2 Quadrature approximation

In this subsection, we focus on evaluating $\mathcal{R}(X)$ using quadrature. Now, recall the mapping $\alpha \mapsto \mathrm{AVaR}_\alpha(X)$ is continuous and increasing on the interval $[0, 1)$ with possible singularity at $\alpha = 1$. With this in mind, we define the constants $0 \leq \beta_- \leq \beta_+ \leq 1$ as

$$\begin{aligned}
\beta_- &:= \sup\{\beta \in [0, 1] \, : \, \mu([0, \beta]) = 0\} \\
\beta_+ &:= \inf\{\beta \in [0, 1] \, : \, \mu([\beta, 1]) = 0\}.
\end{aligned}$$

When $\beta_+ < 1$, $\alpha \mapsto \text{AVaR}_\alpha(X)$ is continuous and increasing on $[\beta_-, \beta_+]$ and hence we can approximate $\mathcal{R}(X)$ using any appropriate quadrature rule for continuous, increasing functions. For example, we could use Gaussian quadrature points and weights corresponding to the system of polynomials that are orthogonal with respect to the weight function, w. However, if $\beta_+ = 1$, then the standard quadrature approximation may not directly apply. We proceed as in the classic works by Pòlya, Davis and Rabinowitz [8,22,23] by applying an $(N+2)$-point quadrature rule to the auxiliary function $g : [0,1] \to \mathbb{R}$ where $g(\alpha) = \text{AVaR}_\alpha(X)$ for all $\alpha < 1$ and $g(1) = 0$. The quadrature rules of interest are defined by the abscissae $\beta_- \le \beta_{N,0} < \beta_{N,1} < \cdots < \beta_{N,N} < \beta_{N,N+1} = 1$ and weights $\{w_{N,n}\}_{n=0}^{N+1}$ satisfying $w_{N,n} > 0$ for all $n = 0, \ldots, N+1$ and $w_{N,0} + \cdots + w_{N,N+1} = 1$. With this scheme, we approximate $\mathcal{R}$ by

$$\mathcal{R}_N(X) := (1 - w_{N,N+1})^{-1} \widehat{\mathcal{R}}_N(X) \tag{17a}$$

where

$$\widehat{\mathcal{R}}_N(X) := \sum_{n=0}^{N} w_{N,n} \text{AVaR}_{\beta_{N,n}}(X). \tag{17b}$$

Note that since $\beta_{N,N+1} = 1$, we do not evaluate $\text{AVaR}_{\beta_{N,N+1}}(X) = \text{AVaR}_1(X)$ as it may be infinite. Additionally, the scaling $(1 - w_{N,N+1})^{-1}$ ensures that $\mathcal{R}_N$ is a convex combination of AVaRs and hence a spectral risk measure.

Although $\alpha \mapsto \text{AVaR}_\alpha(X)$ is continuous and increasing, the "Fundamental Lemma" in [23] does not directly apply since we are integrating with respect to the general measure μ. However, the corollary following the "Fundamental Lemma" in [23] does apply. In this corollary, Rabinowitz treats the weight function, w, as part of the integrand and thus relies on point evaluations of w. This classic result holds for the specific weight functions w resulting in Romberg, Gauss-Legendre, Gauss-Jacobi and Fejér quadrature rules [14]. The following lemma generalizes the "Fundamental Lemma" in [23, pg. 194] to account for integration with respect to the arbitrary measure μ. In addition, we relax the assumption of continuity in [23, pg. 194] and consider quadrature rules that converge for all increasing (possibly discontinuous) functions [16,21]. For this result we assume $\beta_- = 0$. Of course, if $\beta_- > 0$, then we can transform the interval $[\beta_-, 1]$ to $[0,1]$ using the transformation $T(\beta) = (\beta - \beta_-)/(1 - \beta_-)$, which induces the transformation of μ into $\overline{\mu} = \mu \circ T^{-1}$.

Lemma 1 *Let $g : [0,1] \to \mathbb{R}$ be increasing with singularity at $\alpha = 1$. Moreover, assume μ is a probability measure on $[0,1]$ that is absolutely continuous with respect to the Lebesgue measure and satisfies*

$$I(g) = \int_0^1 g(\beta)\, d\mu(\beta) = \lim_{b \to 1^-} \int_0^b g(\beta)\, d\mu(\beta) < \infty.$$

Define the $(N+2)$-point quadrature approximation operator

$$Q_N(f) = \sum_{n=0}^{N+1} w_{N,n} f(\beta_{N,n})$$

with abscissae $0 \leq \beta_{N,0} < \beta_{N,1} < \cdots < \beta_{N,N} < \beta_{N,N+1} = 1$ *and weights satisfying*

$$w_{N,n} > 0 \quad \text{for} \quad n = 0, \ldots, N+1, \quad w_{N,0} + \cdots + w_{N,N+1} = 1.$$

Then, the quadrature approximation

$$\widehat{Q}_N(g) = \sum_{n=0}^{N} w_{N,n} g(\beta_{N,n})$$

converges to $I(g)$ *as* $N \to \infty$ *if:*

1. $Q_N(f)$ *converges to* $I(f)$ *for all finite, increasing functions* $f : [0,1] \to \mathbb{R}$*;*
2. *There exists* $K > 0$ *such that*

$$w_{N,n} \leq K \mu([\beta_{N,n}, \beta_{N,n+1}])$$

for all sufficiently large N *and all* $n \in \{0, \ldots, N+1\}$ *satisfying* $1 - \delta < \beta_{N,n}$ *for some fixed* $\delta > 0$.

Proof Define $f(\alpha) = g(\alpha) - g(0) \geq 0$ for all $\alpha \in [0,1)$. Then it is straight forward to modify the proof of the "Fundamental Lemma" in [23, pg. 194–195], accounting for μ, in order to prove the desired result. □

Under the assumptions of Lemma 1, we have that $\widehat{\mathcal{R}}_N(X)$ converges to $\mathcal{R}(X)$ for any $X \in \mathcal{X}$. Therefore, if $(1 - w_{N,N+1})^{-1}$ converges to 1 as $N \to \infty$, we also have that $\mathcal{R}_N(X)$ converges to $\mathcal{R}(X)$ for any $X \in \mathcal{X}$. The goal now is to construct a quadrature rule that satisfies the assumptions of Lemma 1.

6 Optimal quadrature: the generalized trapezoidal rule

As required by Lemma 1, we must determine a quadrature rule that converges for all finite, increasing functions. Without loss of generality, we will construct a $(N+2)$-point quadrature rule with $\beta_{N,0} = 0$ and $\beta_{N,N+1} = 1$ that converges for all functions in the class $F := \{f : [0,1] \to [0,1] : f \text{ increasing}, \ f(0) = 0, \ f(1) = 1\}$ by minimizing the worst-case error

$$E_N := \sup_{f \in F} |Q_N(f) - I(f)|.$$

The following result is a simple generalization of results in [16, Sect. 5.A] and [21], and demonstrates that the optimal quadrature for the class F is a generalized trapezoidal rule.

Proposition 3 *Let the abscissae* $0 = \beta_{N,0} < \cdots < \beta_{N,N+1} = 1$ *satisfy* $\mu([\beta_{N,n}, \beta_{N,n+1}]) = (N+1)^{-1}$ *for* $n = 0, 1, \ldots, N$ *and define the weights*

$$w_{N,0} = w_{N,N+1} = \frac{1}{2(N+1)} \quad \text{and} \quad w_{N,n} = \frac{1}{N+1} \quad \text{for} \quad n = 1, \ldots, N. \tag{18}$$

Then, the associated $(N+2)$*-point quadrature rule minimizes* E_N *and yields the minimal error* $E_N = (2N+2)^{-1}$.

Proof Let $f \in F$ be arbitrary. Since f is increasing, we have $\int_{\alpha_0}^{\alpha_1} f(\beta)\, \mathrm{d}\mu(\beta) \leq \mu([\alpha_0, \alpha_1]) f(\alpha_1)$ for all $0 \leq \alpha_0 < \alpha_1 \leq 1$. Therefore,

$$I(f) - Q_N(f) \leq \sum_{n=1}^{N+1} |\mu([\beta_{N,n-1}, \beta_{N,n}]) - w_{N,n}| f(\beta_{N,n}). \tag{19}$$

Similarly, we have $\int_{\alpha_0}^{\alpha_1} f(\beta)\, \mathrm{d}\mu(\beta) \geq \mu([\alpha_0, \alpha_1]) f(\alpha_0)$ for any $0 \leq \alpha_0 < \alpha_1 \leq 1$ and thus

$$Q_N(f) - I(f) \leq w_{N,N+1} + \sum_{n=1}^{N} |\mu([\beta_{N,n}, \beta_{N,n+1}]) - w_{N,n}| f(\beta_{N,n}). \tag{20}$$

By minimizing the upper bound on $|Q_N(f) - I(f)|$ generated by (19) and (20) first for the weights $w_{N,n}$ and then for the abscissae $\beta_{N,n}$, we obtain the quadrature rule stated in the proposition. Substituting these $\beta_{N,n}$ and $w_{N,n}$, $n = 0, \ldots, N+1$, into (19) and (20) yields the bounds

$$0 \leq Q_N(f) - I(f) \leq \frac{1}{2(N+1)} \quad \forall\, f \in F.$$

Therefore, $E_N \leq (2N+2)^{-1}$. Now, by setting $f(\alpha) = 1$ for $\alpha \in (0, 1]$ and $f(0) = 0$, we have that $f \in F$ and $|I(f) - Q_N(f)| = (2N+2)^{-1}$. Hence, $E_N = (2N+2)^{-1}$ as desired and the resulting quadrature rule is optimal. □

Remark 3 Of course if $\beta_- > 0$ or $\beta_+ < 1$, then $\mu([\beta_-, \beta_+]) = 1$ and we can transform the interval $[\beta_-, \beta_+]$ to $[0, 1]$ using the transformation $T(\beta) = (\beta - \beta_-)/(\beta_+ - \beta_-)$. This induces a transformation of μ resulting in the measure $\overline{\mu} = \mu \circ T^{-1}$. Applying Proposition 3 to $\overline{\mu}$ and then transforming the resulting quadrature points to $[\beta_-, \beta_+]$ gives the quadrature rule for μ.

The quadrature rule in Proposition 3 converges for all $f \in F$ and thus converges for all finite, increasing functions. Moreover, we have that $w_{N,n} \leq (N+1)^{-1} = \mu([\beta_{N,n}, \beta_{N,n+1}])$ for all $n = 0, \ldots, N$. Therefore, Lemma 1 holds and applying these quadrature abscissae and weights in (17) provides a convergent approximation $\widehat{\mathcal{R}}_N(X)$. Additionally, since $w_{N,N+1} = (2N+2)^{-1} \to 0$ as $N \to \infty$, we have that $\mathcal{R}_N(X)$ converges to $\mathcal{R}(X)$ for any $X \in \mathcal{X}$. The following proposition provides a rate of convergence.

Proposition 4 *Fix $X \in \mathcal{X}$ and let $\widehat{\mathcal{R}}_N$ be defined as in (17) with the quadrature rule in Proposition 3 (transformed to $[\beta_-, 1]$). Moreover, define*

$$\varepsilon_{p,N} := \int_{\beta_{N,N}}^{1} \frac{1}{(1-\beta)^{\frac{1}{p}}}\, d\mu(\beta) \quad \text{and} \quad \Sigma_p(X) := \mathbb{E}[|X - AVaR_{\beta_-}(X)|^p]^{\frac{1}{p}}.$$

Then the following bounds hold

$$0 \leq \mathcal{R}(X) - \widehat{\mathcal{R}}_N(X) \leq \frac{AVaR_{\beta_-}(X)}{2(N+1)} + \varepsilon_{p,N}\, \Sigma_p(X).$$

Proof Similar to the proof of Proposition 1, we can bound

$$\mathrm{AVaR}_\beta(X - \mathrm{AVaR}_{\beta_-}(X)) \leq (1-\beta)^{-\frac{1}{p}}\, \Sigma_p(X) \tag{21}$$

for any $\beta \in (\beta_-, 1)$. Now, to prove the lower bound, the monotonicity of $\alpha \mapsto \mathrm{AVaR}_\alpha(X)$ and the specific form of the quadrature ensure

$$\begin{aligned}
\widehat{\mathcal{R}}_N(X) &\leq \sum_{n=0}^{N} \mu([\beta_{N,n}, \beta_{N,n+1}])\mathrm{AVaR}_{\beta_{N,n}}(X) \\
&\leq \sum_{n=0}^{N} \int_{\beta_{N,n}}^{\beta_{N,n+1}} \mathrm{AVaR}_\beta(X)\, \mathrm{d}\mu(\beta) = \mathcal{R}(X).
\end{aligned}$$

Similarly, for the upper bound we have

$$\begin{aligned}
\mathcal{R}(X) &\leq \sum_{n=1}^{N} \mu([\beta_{N,n-1}, \beta_{N,n}])\mathrm{AVaR}_{\beta_{N,n}}(X) + \int_{\beta_{N,N}}^{1} \mathrm{AVaR}_\beta(X)\, \mathrm{d}\mu(\beta) \\
&= \sum_{n=1}^{N} w_{N,n}\mathrm{AVaR}_{\beta_{N,n}}(X) + \int_{\beta_{N,N}}^{1} \mathrm{AVaR}_\beta(X)\, \mathrm{d}\mu(\beta) \\
&= \widehat{\mathcal{R}}_N(X) + \int_{\beta_{N,N}}^{1} \mathrm{AVaR}_\beta(X)\, \mathrm{d}\mu(\beta) - \frac{\mathrm{AVaR}_{\beta_-}(X)}{2(N+1)}.
\end{aligned}$$

Adding and subtracting $\mathrm{AVaR}_{\beta_-}(X)/(2N+2)$ from the right hand side and using the translation equivariance of AVaR_β gives

$$\mathcal{R}(X) - \widehat{\mathcal{R}}_N(X) \leq \int_{\beta_{N,N}}^{1} \mathrm{AVaR}_\beta(X - \mathrm{AVaR}_{\beta_-}(X))\, \mathrm{d}\mu(\beta) + \frac{\mathrm{AVaR}_{\beta_-}(X)}{2(N+1)}.$$

The desired upper bound follows from this and (21). □

Remark 4 Since Proposition 3 holds for all finite monotonic functions, we can also apply the resulting optimal quadrature rule to approximate the integral of $\alpha \mapsto (\text{AVaR}_\alpha(X))_+$ with respect to any measure ν that is absolutely continuous with respect to the Lebesgue measure and satisfies Assumption 1 (of course, ν must be normalized). Therefore, we can use the generalized trapezoidal rule to approximate the regret and error measures, $\mathcal{V}_\nu$ and $\mathcal{E}_\nu$, respectively. Moreover, Lemma 1 ensures that this approximation converges as $N \to \infty$, providing a viable approximation for superquantile regression.

In addition to the pointwise convergence of $\mathcal{R}_N$ to $\mathcal{R}$, we have the following relation between subgradients of $\widehat{\mathcal{R}}_N$ and subgradients of $\mathcal{R}$.

Proposition 5 *Suppose $\widehat{\mathcal{R}}_N$ in (17) is defined using the quadrature points and weights in Proposition 3 (transformed to $[\beta_-, 1]$). Then, for fixed $X \in \mathcal{X}$ and for any $\epsilon > 0$ there exists $\overline{N} = \overline{N}(X, \epsilon) \in \mathbb{N}$ satisfying for any $\theta_N \in \partial\widehat{\mathcal{R}}_N(X)$ with $N \geq \overline{N}$ and any $\theta_- \in \partial AVaR_{\beta_-}(X)$,*

$$\theta_N + \frac{1}{2(N+1)}\theta_- \in \partial_\epsilon \mathcal{R}(X) := \{\theta \in \mathcal{X}^* : \mathcal{R}(Y) - \mathcal{R}(X) \geq \mathbb{E}[\theta(Y - X)] - \epsilon\}.$$

Proof Fix $X \in \mathcal{X}$ and define $\bar{X} = X - \text{AVaR}_{\beta_-}(X)$ and $\bar{Y} = Y - \text{AVaR}_{\beta_-}(Y)$ for any $Y \in \mathcal{X}$. By Proposition 4, we have that

$$\mathcal{R}(\bar{Y}) - \mathcal{R}(\bar{X}) \geq \widehat{\mathcal{R}}_N(\bar{Y}) - \widehat{\mathcal{R}}_N(\bar{X}) - \varepsilon_{p,N}\, \Sigma_p(\bar{X}) \quad \forall\, Y \in \mathcal{X}$$

where $\varepsilon_{p,N}$ and $\Sigma_p(\bar{X}) = \Sigma_p(X)$ are defined in Proposition 4. The translation equivariance of $\mathcal{R}$ and the fact that

$$\widehat{\mathcal{R}}_N(X' + c) = \widehat{\mathcal{R}}_N(X') + c \sum_{n=0}^{N} w_{N,n} = \widehat{\mathcal{R}}_N(X') + c(1 - w_{N,N+1})$$

for any $X' \in \mathcal{X}$ and any $c \in \mathbb{R}$ yields

$$\begin{aligned}\mathcal{R}(Y) - \mathcal{R}(X) \geq \widehat{\mathcal{R}}_N(Y) - \widehat{\mathcal{R}}_N(X) - \varepsilon_{p,N}\, \Sigma_p(X) \\ + w_{N,N+1}(\text{AVaR}_{\beta_-}(Y) - \text{AVaR}_{\beta_-}(X)).\end{aligned} \tag{22}$$

Since $\varepsilon_{p,N} \to 0$ as $N \to \infty$ and $\Sigma_p(X) < \infty$, for any $\epsilon > 0$ there exists $\overline{N} \in \mathbb{N}$ such that $\varepsilon_{p,N}\, \Sigma_p(X) < \epsilon$ when $N > \overline{N}$. Thus, using the subgradient inequality for $\widehat{\mathcal{R}}_N$ and AVaR_{β_-} gives

$$\mathcal{R}(Y) - \mathcal{R}(X) \geq \mathbb{E}[(\theta_N + w_{N,N+1}\theta_-)(Y - X)] - \epsilon$$

for any $\theta_N \in \partial\widehat{\mathcal{R}}_N(X)$ and $\theta_- \in \partial\text{AVaR}_{\beta_-}(X)$ proving the desired result. □

In the following proposition, we investigate the limiting properties of the subdifferential $\partial\widehat{\mathcal{R}}_N(X_N)$ where $X_N \to X$ in $\mathcal{X}$. For this result, we use the notation

$$\sigma - \operatorname*{Lim\,Sup}_{N\to\infty} A_N,$$

for $A_N \subseteq \mathcal{X}^*$ with $N = 1, 2, \ldots$, to denote the *sequential weak* upper limit* of the sequence of sets A_N. That is, the set of weak* cluster points of sequences $\{\theta_N\} \subset \mathcal{X}^*$ with $\theta_N \in A_N$.

Proposition 6 *Let $\{X_N\} \subset \mathcal{X}$ with $X_N \to X$ in $\mathcal{X}$. If the assumptions of Proposition 5 hold, then*

$$\emptyset \neq \sigma - \operatorname*{LimSup}_{N\to\infty} \partial\widehat{\mathcal{R}}_N(X_N) \subseteq \partial\mathcal{R}(X).$$

Proof Let $\{\theta_N\} \subset \mathcal{X}^*$ with $\theta_N \in \partial\widehat{\mathcal{R}}_N(X_N)$ be arbitrary. The convexity of AVaR_β and Corollary 3 of Propositions 2.3.3 in [6] ensure

$$\partial\widehat{\mathcal{R}}_N(Y) = w_{N,0}\partial\mathrm{AVaR}_{\beta_{N,0}}(Y) + \cdots + w_{N,N}\partial\mathrm{AVaR}_{\beta_{N,N}}(Y)$$

for any $Y \in \mathcal{X}$ and any $N \in \mathbb{N}$. Therefore, for any $\theta_N \in \partial\widehat{\mathcal{R}}_N(X)$ there exists $\theta_{N,n} \in \partial\mathrm{AVaR}_{\beta_{N,n}}(X)$ with $n = 0, \ldots, N$ satisfying $\theta_N = w_{N,0}\theta_{N,0} + \cdots + w_{N,N}\theta_{N,N}$ and thus

$$\|\theta_N\|_{\mathcal{X}^*} \leq w_{N,0}\|\theta_{N,0}\|_{\mathcal{X}^*} + \cdots + w_{N,N}\|\theta_{N,N}\|_{\mathcal{X}^*}.$$

The explicit form of the subdifferential of $\mathrm{AVaR}_\beta(X)$ with $\beta \in [0, 1)$ (see, e.g., [31]) combined with Hölder's inequality ensures $\|\theta_{N,n}\|_{\mathcal{X}^*} \leq 1/(1-\beta_{N,n})^{\frac{1}{p}}$ for all $n = 0, \ldots, N$ and hence

$$\|\theta_N\|_{\mathcal{X}^*} \leq \sum_{n=0}^{N} \frac{w_{N,n}}{(1-\beta_{N,n})^{\frac{1}{p}}} \leq \int_0^1 \frac{1}{(1-\beta)^{\frac{1}{p}}}\,\mathrm{d}\mu(\beta) < \infty \quad \forall\, N.$$

Thus, Alaoglu's Theorem [11, Thm. 5.18] ensures the existence of a weak* converging subsequence $\{\theta_{N_k}\}$ with weak* limit $\theta \in \mathcal{X}^*$.

Now, by adding and subtracting terms, we arrive at the following equality,

$$\begin{aligned}\mathcal{R}(Y) - \mathcal{R}(X) = {} & (\mathcal{R}(Y) - \widehat{\mathcal{R}}_{N_k}(Y)) + (\widehat{\mathcal{R}}_{N_k}(Y) - \widehat{\mathcal{R}}_{N_k}(X_{N_k})) \\ & + (\widehat{\mathcal{R}}_{N_k}(X_{N_k}) - \mathcal{R}(X_{N_k})) + (\mathcal{R}(X_{N_k}) - \mathcal{R}(X))\end{aligned}$$

for all $Y \in \mathcal{X}$. The first and last terms on the right-hand side have limit zero by Lemma 1 and [31, Prop. 6.6] (i.e., $\mathcal{R}$ is coherent and hence continuous), respectively. Proposition 4 and the continuity of the maps $Y \mapsto \mathrm{AVaR}_\alpha(Y)$ for $\alpha \in [0, 1)$ and

$Y \mapsto \Sigma_p(Y)$ ensure that the third term also has limit zero. Finally, the second term is bounded below as

$$\widehat{\mathcal{R}}_{N_k}(Y) - \widehat{\mathcal{R}}_{N_k}(X_{N_k}) \geq \mathbb{E}[\theta_{N_k}(Y - X_{N_k})] = \mathbb{E}[\theta_{N_k}(Y - X)] + \mathbb{E}[\theta_{N_k}(X - X_{N_k})]$$

and hence the weak* convergence of θ_{N_k} ensures

$$\mathcal{R}(Y) - \mathcal{R}(X) \geq \mathbb{E}[\theta(Y - X)] + \lim_{N_k \to \infty} \mathbb{E}[\theta_{N_k}(X - X_{N_k})] \quad \forall\, Y \in \mathcal{X}.$$

Since $\{\theta_{N_k}\}$ is bounded and $X_{N_k} \to X$, we have that $\mathbb{E}[\theta_{N_k}(X - X_{N_k})]$ converges to zero which implies $\theta \in \partial\mathcal{R}(X)$. □

Corollary 1 *Let the assumptions of Proposition 6 hold. Then,*

$$\emptyset \neq \sigma - \text{Limsup}_{N \to \infty}\, \partial\mathcal{R}_N(X_N) \subseteq \partial\mathcal{R}(X).$$

Proof Recall that $\mathcal{R}_N = (1 - w_{N,N+1})^{-1}\widehat{\mathcal{R}}_N$. Therefore,

$$\partial\mathcal{R}_N(Y) = (1 - w_{N,N+1})^{-1}\partial\widehat{\mathcal{R}}_N(Y)$$

for any $Y \in \mathcal{X}$. Since $(1 - w_{N,N+1})^{-1} \in [1, 4/3]$ with $(1 - w_{N,N+1})^{-1} \downarrow 1$, the result follows from Proposition 6. □

7 Optimization

To conclude our analysis, we investigate the behavior of our target optimization problem (1) when $\mathcal{R}$ is replaced by the quadrature approximation $\mathcal{R}_N$ in (17). The goal of this section is to prove consistency of this approximation, but first we prove existence of minimizers.

Proposition 7 *Suppose $\mathcal{H} : \mathcal{X} \to \mathbb{R}$ is a coherent risk measure. Additionally, let Z be a reflexive Banach space and let $Z_{ad} \subset Z$ be a convex, closed and bounded set. Finally, assume $J : Z \to \mathcal{X}$ is completely continuous, i.e.,*

$$z_n \rightharpoonup z \text{ in } Z \implies J(z_n) \to J(z) \text{ in } \mathcal{X}.$$

Then, there exists a solution to the optimization problem

$$\min_{z \in Z_{ad}} \mathcal{H}(J(z)).$$

Namely, (1) has a solution.

Proof The complete continuity of J and the continuity of $\mathcal{H}$ (i.e., Proposition 6.6 in [31]) ensure that $\mathcal{H} \circ J$ is weakly continuous and hence weakly lower semicontinuous. Therefore, the result follows from the direct method of the calculus of variations, e.g., Corollary 3.2.3 in [4]. □

Remark 5 We can similarly show existence of solutions when Z_{ad} is unbounded if, e.g., $\mathcal{H} \circ J$ is coercive.

Proposition 7 guarantees existence of minimizers for $\mathcal{R} \circ J$ and $\mathcal{R}_N \circ J$ over Z_{ad} for all $N = 1, 2, \ldots$. Additionally, Proposition 3 combined with Lemma 1 ensures that $\mathcal{R}_N(X)$ converges to $\mathcal{R}(X)$ as N goes to infinity for any $X \in \mathcal{X}$. These results are pivotal in proving consistency of our approximation.

Theorem 3 *Let the assumptions of Proposition 7 hold and let μ in the definition of $\mathcal{R}$, be a probability measure on* $[0, 1]$ *that is absolutely continuous with respect to the Lebesgue measure and satisfies Assumption* 1.3. *Moreover, assume that the quadrature rule defining $\mathcal{R}_N$ is given by Proposition 3 and let $z_N \in Z_{ad}$ denote a minimizer of $\mathcal{R}_N \circ J$ over Z_{ad} for $N = 1, 2, \ldots$. Then, any weak accumulation point of $\{z_N\}$ minimizes $\mathcal{R} \circ J$ over Z_{ad}.*

Proof First, the assumptions on μ and Proposition 7 ensure that z_N exists for each N and $\mathcal{R} \circ J$ has a minimizer in Z_{ad}. Moreover, $\{z_N\}$ is bounded and hence has a weak accumulation point, denoted $z^\star \in Z_{\text{ad}}$ with associated weakly converging subsequence $z_{N_k} \rightharpoonup z^\star$. By Lemma 1 and the optimality of z_N, we have that

$$\mathcal{R}(J(z)) = \lim_{N\to\infty} \mathcal{R}_N(J(z)) \geq \liminf_{N\to\infty} \mathcal{R}_N(J(z_N)) \geq \liminf_{N\to\infty} \widehat{\mathcal{R}}_N(J(z_N)) \tag{23}$$

for all $z \in Z_{\text{ad}}$. Moreover, Proposition 4 ensures that for any $Y \in \mathcal{X}$, we have

$$\mathcal{R}(Y) - \widehat{\mathcal{R}}_N(Y) \leq \frac{\text{AVaR}_{\beta_-}(Y)}{2(N+1)} + \varepsilon_{p,N}\, \Sigma_p(Y).$$

Additionally, $Y \mapsto \Sigma_p(Y)$ is continuous (since it is the composition of the L^p-norm with the continuous function $Y \mapsto (Y - \text{AVaR}_\alpha(Y))$ and therefore the complete continuity of J ensures that $\text{AVaR}_{\beta_-}(J(z_{N_k}))/(2N_k + 2) \to 0$ and $\varepsilon_{p,N_k}\, \Sigma_p(J(z_{N_k})) \to 0$ as $N_k \to \infty$. Hence,

$$\begin{aligned} \mathcal{R}(J(z^\star)) &\leq \liminf_{N_k\to\infty} \left\{ \widehat{\mathcal{R}}_{N_k}(J(z_{N_k})) + \frac{\text{AVaR}_{\beta_-}(J(z_{N_k}))}{2(N_k+1)} + \varepsilon_{p,N_k}\, \Sigma_p(J(z_{N_k})) \right\} \\ &= \liminf_{N_k\to\infty} \left\{ \widehat{\mathcal{R}}_{N_k}(J(z_{N_k})) \right\}. \end{aligned} \tag{24}$$

Therefore, (23) and (24) guarantee that $\mathcal{R}(J(z^\star)) \leq \mathcal{R}(J(z))$ for all $z \in Z_{\text{ad}}$ which proves the desired result. □

Theorem 3 demonstrates consistency of our approximation for minimizers of the composite functional $\mathcal{J}(z) = \mathcal{R}(J(z))$. Unfortunately, if $\mathcal{J}$ is not convex, then numerical methods typically can only guarantee first-order stationary points. For our final result, we prove consistency of our approximation for stationary points of $\mathcal{J}$. In this context, a stationary point of $\mathcal{J}$ over Z_{ad} is a vector $z \in Z_{\text{ad}}$ satisfying

$$0 \in \partial_C \mathcal{J}(z) + \mathcal{N}_{Z_{\text{ad}}}(z)$$

where $\partial_C \mathcal{J}(z)$ denotes the Clarke subdifferential of $\mathcal{J}$ and $\mathcal{N}_{Z_{\rm ad}}(z)$ denotes the normal cone of the convex set $Z_{\rm ad}$ at z. We analogously define stationary points of $\mathcal{J}_N = \mathcal{R}_N \circ J$ over $Z_{\rm ad}$. Note that if J is continuously Fréchet differentiable, then by Theorem 2.3.10 in [6], we have

$$\partial_C \mathcal{J}(z) = \nabla J(z)^* \partial \mathcal{R}(J(z))$$

(i.e., the composition is regular since $\mathcal{R}$ is convex). If $\nabla J(z) \in L^p(\Omega, \mathcal{F}, \mathbb{P}; Z^*)$ where $L^p(\Omega, \mathcal{F}, \mathbb{P}; Z^*)$ denotes the Banach space of $\mathcal{F}$-strongly measurable functions $g : \Omega \to Z^*$ that satisfy $\mathbb{E}[\|g\|_{Z^*}^p] < \infty$ (i.e., a Bochner space), then

$$\eta \in \partial_C \mathcal{J}(z) \quad \Longleftrightarrow \quad \exists\, \theta \in \partial \mathcal{R}(J(z)) \quad \text{such that} \quad \eta = \mathbb{E}[\theta \nabla J(z)].$$

Similar results hold for $\mathcal{J}_N$. Note that $\nabla J(z)$ is only required to be a bounded linear operator mapping Z into $\mathcal{X}$. By defining $g(z, \omega)v := [\nabla J(z)v](\omega)$ for all $v \in Z$, we therefore have that $g(z, \cdot)$ is a strongly $\mathcal{F}$-measurable operator-valued function [15, Def. 3.5.5(2)] and $|g(z, \cdot)v|^p$ is integrable for all $z,\ v \in Z$. Thus, the assumption that $g(z, \cdot) \in L^p(\Omega, \mathcal{F}, \mathbb{P}; Z^*)$ imposes the additional constraints that $g(z, \cdot)$ is a uniformly $\mathcal{F}$-measurable operator-valued function [15, Def. 3.5.5(i)] and $\|g(z, \cdot)\|_{Z^*}^p$ is integrable for all $z \in Z$. We use the above characterization of subgradients to prove the following result.

Theorem 4 *Let the assumptions of Theorem 3 hold. Further assume that J is continuously Fréchet differentiable and its derivative, $\nabla J : Z \to L^p(\Omega, \mathcal{F}, \mathbb{P}; Z^*)$, is completely continuous, i.e.,*

$$z_n \rightharpoonup z \text{ in } Z \implies \nabla J(z_n) \to \nabla J(z) \text{ in } L^p(\Omega, \mathcal{F}, \mathbb{P}; Z^*).$$

Finally, let $z_N \in Z_{ad}$ denote stationary points of $\mathcal{J}_N := \mathcal{R}_N \circ J$ over Z_{ad} for $N = 1, 2, \ldots$. Then, any weak accumulation point of $\{z_N\}$ is a stationary point of $\mathcal{J} := \mathcal{R} \circ J$ over Z_{ad}.

Proof Since $Z_{\rm ad}$ is bounded, the sequence of stationary points $\{z_N\}$ has a weakly converging subsequence $\{z_{N_k}\}$ with weak limit $z^\star \in Z_{\rm ad}$. Moreover, since $\{z_{N_k}\} \subset Z_{\rm ad}$ are stationary points, we have that

$$\exists\, \eta_{N_k} \in \partial_C \mathcal{J}_{N_k}(z_{N_k}) \quad \text{with} \quad -\eta_{N_k} \in \mathcal{N}_{Z_{\rm ad}}(z_{N_k}).$$

Associated with each η_{N_k} is a subgradient $\theta_{N_k} \in \partial \mathcal{R}_{N_k}(J(z_{N_k}))$ satisfying $\eta_{N_k} = \mathbb{E}[\theta_{N_k} \nabla J(z_{N_k})]$. As shown in Proposition 6, $\{\theta_{N_k}\}$ is a bounded sequence and hence has a weak* converging subsequence $\{\theta_{N_{k_j}}\}$ with weak* limit $\theta^\star \in \partial \mathcal{R}(J(z^\star))$. Therefore, $\mathbb{E}[\theta^\star \nabla J(z^\star)] \in \partial \mathcal{J}(z^\star)$. This and the complete continuity of ∇J ensure that

$$\mathbb{E}[\theta_{N_{k_j}} \nabla J(z_{N_{k_j}})] \rightharpoonup \mathbb{E}[\theta^\star \nabla J(z^\star)] \quad \text{in } Z^*.$$

It remains to show that $-\mathbb{E}[\theta^\star \nabla J(z^\star)] \in \mathcal{N}_{Z_{\rm ad}}(z^\star)$. Since $z_{N_{k_j}}$ is a stationary point, we have

$$0 \leq \langle \eta_{N_{k_j}}, z - z_{N_{k_j}} \rangle_{Z^*,Z} = \mathbb{E}[\theta_{N_{k_j}} \langle \nabla J(z^\star), z - z_{N_{k_j}} \rangle_{Z^*,Z}] + \mathbb{E}[\theta_{N_{k_j}} \langle \nabla J(z_{N_{k_j}}) - \nabla J(z^\star), z - z_{N_{k_j}} \rangle_{Z^*,Z}] \quad (25)$$

for all $z \in Z_{\rm ad}$. Fix an arbitrary $z \in Z_{\rm ad}$. To handle the first term on the right-hand side of (25), we note that $\{z_{N_{k_j}}\}$ converges weakly and therefore is bounded. This implies the existence of $M \geq 0$ such that

$$|\langle \nabla J(z^\star), z - z_{N_{k_j}} \rangle_{Z^*,Z}| \leq M \|\nabla J(z^\star)\|_{Z^*} \quad \mathbb{P}\text{-a.e.}$$

for all N_{k_j}. In addition, the weak convergence of $\{z_{N_{k_j}}\}$ implies

$$\langle \nabla J(z^\star), z - z_{N_{k_j}} \rangle_{Z^*,Z} \to \langle \nabla J(z^\star), z - z^\star \rangle_{Z^*,Z} \quad \mathbb{P}\text{-a.e.} \quad (26)$$

Thus the Lebesgue Dominated Convergence Theorem [11, Th. 2.24] ensures strong convergence of (26) in $\mathcal{X}$ which, combined with the weak* convergence of $\{\theta_{N_{k_j}}\}$, guarantees

$$\mathbb{E}[\theta_{N_{k_j}} \langle \nabla J(z^\star), z - z_{N_{k_j}} \rangle_{Z^*,Z}] \to \mathbb{E}[\theta^\star \langle \nabla J(z^\star), z - z^\star \rangle_{Z^*,Z}].$$

Additionally, we bound the second term on the right-hand side of (25) using

$$\|\langle \nabla J(z_{N_{k_j}}) - \nabla J(z^\star), z - z_{N_{k_j}} \rangle_{Z^*,Z}\|_{\mathcal{X}} \leq M \|\nabla J(z_{N_{k_j}}) - \nabla J(z^\star)\|_{L^p(\Omega,\mathcal{F},\mathbb{P};Z^*)}$$

and note that the right-hand side of this bound converges to zero due to the complete continuity of ∇J. This, along with the weak* convergence of $\{\theta_{N_{k_j}}\}$, guarantees

$$\mathbb{E}[\theta_{N_{k_j}} \langle \nabla J(z_{N_{k_j}}) - \nabla J(z^\star), z - z_{N_{k_j}} \rangle_{Z^*,Z}] \to 0.$$

Therefore, since $z \in Z_{\rm ad}$ is arbitrary, passing to the limit in (25) gives the desired result. □

8 Numerical results

8.1 Quadrature approximation error

We now demonstrate the error estimates in Propositions 3 and 4 on two examples: (i) a random variable X that is uniformly distributed on [0, 1] and (ii) a random variable X that is exponentially distributed with shape parameter $\lambda = 1$. For both examples,

Table 1 Quantile function, average value-at-risk and spectral risk for examples (i) and (ii)

Ex.	$F_X^{-1}(\beta)$	$\text{AVaR}_\beta(X)$	$\mathcal{R}(X)$
(i)	β	$\frac{1}{2}(1+\beta)$	$\frac{1}{2}(1+p_1/(p_1+p_2))$
(ii)	$-\log(1-\beta)$	$(1-\log(1-\beta))$	$(1-\psi(p_2)+\psi(p_1+p_2))$

ψ denotes the digamma function and p_1, p_2 are the shape parameters of the beta distribution

we set μ to be the measure associated with the beta distribution with various shape parameters p_1 and p_2, i.e., μ has density function

$$w(\beta) = \frac{\beta^{p_1-1}(1-\beta)^{p_2-1}}{B(p_1, p_2)} \quad \text{where} \quad B(p_1, p_2) := \frac{\Gamma(p_1)\Gamma(p_2)}{\Gamma(p_1+p_2)}.$$

The quantile functions, the AVaRs and the spectral risks for these examples are listed in Table 1. Table 2 lists the error associated with $\widehat{\mathcal{R}}_N(X)$. The row titled "Optimal" lists the error bound from Proposition 3. The left and center columns of Fig. 3 depict the errors from Table 2 on a log-log scale for various shape parameter configurations. The blue lines depict the errors associated with $\widehat{\mathcal{R}}_N(X)$ whereas the red lines depict the errors associated with $\mathcal{R}_N(X)$. The black asterisks denote the theoretical error bound from Proposition 3. Moreover, the right column of Fig. 3 depicts the quadrature abscissae (red asterisks) associated with the 12 point quadrature approximation ($N = 10$). The probability density functions (blue lines) are plotted above the abscissae to help visualize the distribution of points. As depicted in Table 2 and Fig. 3, the quadrature errors converge at nearly the optimal rate even though they are only guaranteed to satisfy the modified bound from Proposition 4. This behavior is likely a consequence of the asymptotic properties of $\varepsilon_{N,p}$ in Proposition 4 for the specified measures μ.

8.2 Optimal contaminant mitigation

In this section, we formulate and solve an idealized optimal control problem representing the mitigation of a pollutant. This problem was first studied in [17]. We model the spread of the pollutant with the steady advection-diffusion equation. The goal is to determine locations in which to inject chemicals that dissolve the pollutant and thus minimize the total concentration. The uncertainties arise in the advection field, the diffusion coefficient, and the sources. Let $D = (0, 1)^2$ denote the physical domain and $U := H^1(D)$ be the space of pollutant concentrations. Here, $H^1(D)$ is the Sobolev space of square integrable functions whose weak derivatives are also square integrable [2], i.e.,

$$H^1(D) := \{u \in L^2(D) \,:\, \nabla u \in L^2(D)\}.$$

Table 2 The error between $\widehat{\mathcal{R}}_N(X)$ and $\mathcal{R}(X)$ for the two example random variables X and various choices of μ

N	10^1	10^2	10^3	10^4	10^5
Uniform					
Optimal	4.545×10^{-2}	4.950×10^{-3}	4.995×10^{-4}	5.000×10^{-5}	5.000×10^{-6}
Beta($\frac{1}{2}, \frac{1}{2}$)	4.545×10^{-2}	4.950×10^{-3}	4.995×10^{-4}	5.000×10^{-5}	5.000×10^{-6}
Beta(1,5)	3.568×10^{-2}	4.264×10^{-3}	4.556×10^{-4}	4.722×10^{-5}	4.825×10^{-6}
Beta(5,1)	5.522×10^{-2}	5.637×10^{-3}	5.434×10^{-4}	5.277×10^{-5}	5.175×10^{-6}
Beta(5,2)	5.201×10^{-2}	5.423×10^{-3}	5.300×10^{-4}	5.193×10^{-5}	5.122×10^{-6}
Beta(2,5)	3.890×10^{-2}	4.478×10^{-3}	4.690×10^{-4}	4.806×10^{-5}	4.878×10^{-6}
Beta(2,2)	4.545×10^{-2}	4.950×10^{-3}	4.995×10^{-4}	5.000×10^{-5}	5.000×10^{-6}
Exponential					
Beta($\frac{1}{2}, \frac{1}{2}$)	3.895×10^{-1}	6.437×10^{-2}	8.786×10^{-3}	1.110×10^{-4}	1.340×10^{-5}
Beta(1,5)	8.410×10^{-2}	1.134×10^{-2}	1.373×10^{-3}	1.605×10^{-4}	1.835×10^{-5}
Beta(5,1)	3.367×10^{-1}	4.647×10^{-2}	5.769×10^{-3}	6.887×10^{-4}	8.016×10^{-5}
Beta(5,2)	2.189×10^{-1}	2.867×10^{-2}	3.425×10^{-3}	3.979×10^{-4}	4.540×10^{-5}
Beta(2,5)	9.896×10^{-2}	1.299×10^{-2}	1.543×10^{-3}	1.778×10^{-4}	2.010×10^{-5}
Beta(2,2)	1.690×10^{-1}	2.372×10^{-2}	2.961×10^{-3}	3.537×10^{-4}	1.743×10^{-5}

The measures μ are chosen to represent beta distributions. The row titled "Optimal" lists the error bound from Proposition 3

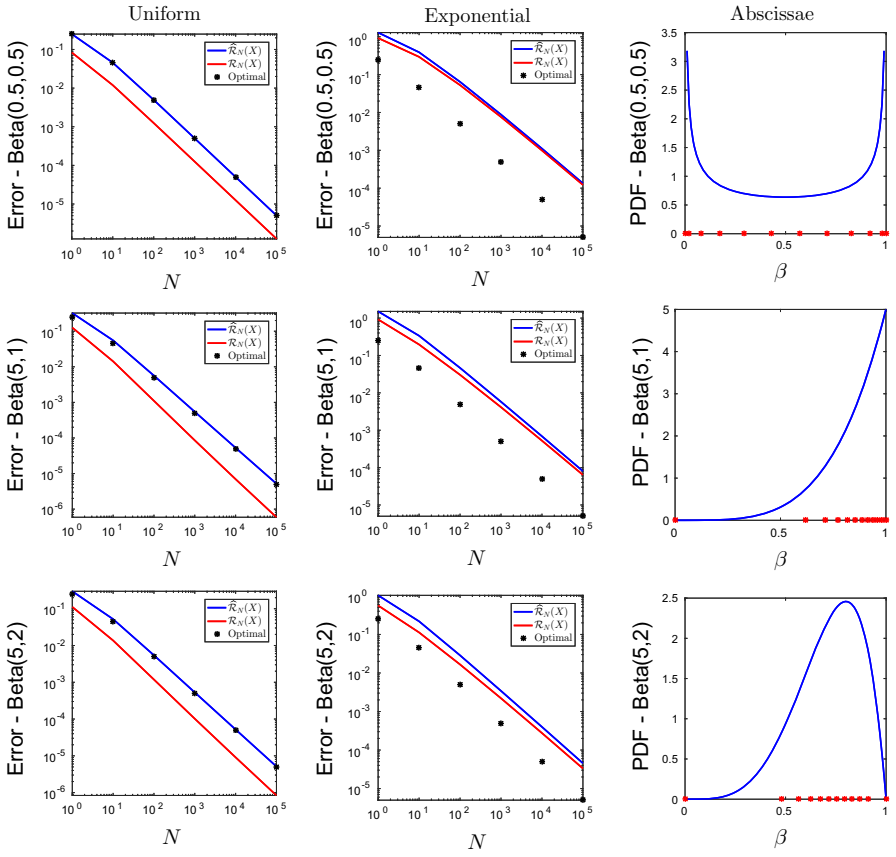

Fig. 3 The error between $\widehat{\mathcal{R}}_N(X)$ and $\mathcal{R}(X)$ (blue line) and between $\mathcal{R}_N(X)$ and $\mathcal{R}(X)$ (red line) for the two example random variables X and various choices of μ. The measures μ are chosen to represent beta distributions. The right images depict the quadrature points (red asterisks) associated with the 12 point quadrature approximation ($N = 10$)

The target optimization problem is

$$\min_{z\in Z_{\text{ad}}} \frac{\kappa_s}{2}\mathcal{R}\left(\int_D S(z)^2\,\mathrm{d}x\right) + \wp(z) \tag{27}$$

where $\kappa_s > 0$ and $S(z) = u : \Omega \to U$ solves

$$\int_D \{\epsilon(\omega)\nabla u\cdot\nabla v + (\mathbb{V}(\omega)\cdot\nabla u)v\}\,\mathrm{d}x = \int_D (f(\omega) - Bz)v\,\mathrm{d}x \quad \forall\, v \in U_d. \tag{28}$$

Here $U_d := \{v \in U \,:\, v|_{\Gamma_d} = 0\}$ with $\Gamma_d := \{0\}\times(0,1)$. We enforce homogeneous Dirichlet boundary conditions on Γ_d and homogeneous Neumann boundary conditions on the remaining boundary. The control space is $Z = \mathbb{R}^9$ with $Z_{\text{ad}} = \{z \in Z \,:\, 0 \le z \le 1\}$ and control cost

$$\wp(z) := \kappa_c \sum_{k=1}^{9} z_k, \quad \kappa_c > 0.$$

For the definitions of the PDE coefficients ϵ, $\mathbb{V}$, f, and B, see [17, Sect. 4.1]. As shown in [17], the PDE (28) has a unique solution $S(z)$ which depends continuously on z. Moreover, $z \mapsto S(z) : Z \to L^s(\Omega, \mathcal{F}, \mathbb{P}; U)$ for any $s \in [1, \infty)$ is continuously Fréchet differentiable. This and the fact that Z is finite-dimensional ensure that the random variable objective function

$$J(z) := \frac{\kappa_s}{2} \int_D S(z)^2 \, \mathrm{d}x + \wp(z)$$

and its gradient ∇J are completely continuous. Thus, Theorems 3 and 4 apply. Furthermore, J is convex since S is a linear operator. Therefore, the stationarity conditions introduced in the previous section are necessary and sufficient for a vector $z \in Z_{\text{ad}}$ to be a minimizer of (27). In particular, these results ensure that the sequence of minimizers of $\mathcal{R}_N \circ J$ over Z_{ad} for increasing N will contain a convergent subsequence whose limit solves (27).

We formulate this optimization problem using the spectral risk measure $\mathcal{R}$ in which μ is defined to be the measure representing the beta distribution with $p_1 = 5$ and $p_2 = 2$. We chose this μ because it places considerable weight on AVaR confidence levels $\alpha \in (0.5, 1)$ but also places nonzero weight on smaller (less conservative) confidence levels. This choice results in conservative optimal solutions that also perform well on average. We approximate $\mathcal{R}$ with the optimal quadrature described above. Moreover, we approximate the expectation in (5) with 1000 Monte Carlo samples. In Table 3, we list the resulting optimal controls, optimal objective function values and optimal statistics for each approximate risk measure $\mathcal{R}_N$ with $N + 1 \in \{2, 4, 8, 16, 32, 64, 128, 256, 512, 1024\}$. In Table 4, we list the change in the optimal control and statistic between consecutive quadrature levels. We see that the differences appear to decrease monotonically, suggesting convergence of the optimal controls and statistics.

9 Conclusions

We have developed a family of risk quadrangles that generates a large class of spectral risk measures. We demonstrate that the statistic associated with these quadrangles is the average value-at-risk (or an interval of average values-at-risk). As a consequence, the error measures associated with this family of quadrangles can be used to perform superquantile regression. It is left as future work to compare the quality of regression using these general error measures with the error measure introduced in [28].

In addition to the quadrangle, we have introduced an optimal quadrature approximation for spectral risk measures in which the defining measure μ is absolutely continuous with respect to the Lebesgue measure. This quadrature approximation is a generalization of the trapezoidal rule for general measures μ. We have shown that this approximation converges for all increasing functions with possible singularity at the

Table 3 Controls, minimal objective value and statistic for varying $\mathcal{R}_N$

$N+1$	1	2	3	4	5	6	7	8	9	obj	stat
2	–	1.000	–	0.188	1.000	0.127	–	–	–	4.593	2.563
4	–	1.000	–	0.321	1.000	0.323	–	–	–	4.925	3.781
8	–	1.000	–	0.383	1.000	0.393	–	–	–	5.076	3.951
16	–	1.000	–	0.413	1.000	0.430	–	–	–	5.154	4.040
32	–	1.000	–	0.427	1.000	0.453	–	–	–	5.196	4.093
64	–	1.000	–	0.435	1.000	0.467	–	–	–	5.219	4.119
128	–	1.000	–	0.439	1.000	0.473	–	–	–	5.232	4.135
256	–	1.000	–	0.441	1.000	0.477	–	–	–	5.240	4.142
512	–	1.000	–	0.442	1.000	0.479	–	–	–	5.245	4.148
1024	–	1.000	–	0.442	1.000	0.481	–	–	–	5.247	4.151

Table 4 The change between the optimal controls and statistics computed using $\mathcal{R}_{N_k}$ and $\mathcal{R}_{N_{k-1}}$ where $N_k = 2^k - 1$ for $k = 2, \ldots, 10$

Change	4	8	16	32	64	128	256	512	1024
Control	0.2361	0.0941	0.0477	0.0267	0.0155	0.0076	0.0041	0.0028	0.0015
Stat	1.2184	0.1696	0.0888	0.0530	0.0267	0.0154	0.0069	0.0060	0.0028

right end point. Thus, one can additionally use this quadrature to approximate the error measures for superquantile regression. Finally, we proved consistency of minimizers and first-order stationary points when using this approximation and demonstrated our results on two numerical examples.

References

1. Acerbi, C.: Spectral measures of risk: a coherent representation of subjective risk aversion. J. Bank. Finance **26**, 1505–1518 (2002)
2. Adams, R.A., Fournier, J.J.F.: Sobolev Spaces, 2nd edn. Elsevier, Amsterdam (2003)
3. Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. Math. Finance **9**, 203–228 (1999)
4. Attouch, H., Buttazzo, G., Michaille, G.: Variational Analysis in Sobolev and BV Spaces. MPS/SIAM Series on Optimization, vol. 6. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (2006)
5. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, Berlin (2000)
6. Clarke, F.H.: Optimization and Nonsmooth Analysis. Society for Industrial and Applied Mathematics, Philadelphia (1990)
7. Conti, S., Held, H., Pach, M., Rumpf, M., Schultz, R.: Risk averse shape optimization. SIAM J. Control Optim. **49**, 927–947 (2011)
8. Davis, P.J., Rabinowitz, P.: Ignoring the singularity in approximate integration. J. Soc. Ind. Appl. Math. Ser. B Numer. Anal. **2**, 367–383 (1965)
9. Denneberg, D.: Distorted probabilities and insurance premiums. Methods Oper. Res. **63**, 3–5 (1990)
10. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems, Classics in Applied Mathematics, vol. 28. SIAM, Philadelphia (1999)

11. Folland, G.B.: Real Analysis. Modern Techniques and Their Applications. Pure and Applied Mathematics (New York), 2nd edn. Wiley, New York (1999)
12. Föllmer, H., Schied, A.: Stochastic Finance: An Introduction in Discrete Time, Vol. 27 of de Gruyter Studies in Mathematics, 3 revised and extended edn. Walter de Gruyter & Co., Berlin (2011)
13. Gaivoronski, A.A., Pflug, G.: Value at risk in portfolio optimization: properties and computational approach. J. Risk **7**, 1–31 (2005)
14. Gautschi, W.: Numerical quadrature in the presence of a singularity. SIAM J. Numer. Anal. **4**, 357–362 (1967)
15. Hille, E., Phillips, R.S.: Functional Analysis and Semi-Groups. American Mathematical Society Colloquium Publications, vol. 31. American Mathematical Society, Providence, rev. ed (1957)
16. Kiefer, J.: Optimum sequential search and approximation methods under minimum regularity assumptions. J. Soc. Ind. Appl. Math. **5**, 105–136 (1957)
17. Kouri, D.P., Surowiec, T.M.: Existence and optimality conditions for risk-averse PDE-constrained optimization. SIAM/ASA J. Uncertain. Quantif. (to appear, 2018)
18. Kouri, D.P., Surowiec, T.M.: Risk-averse PDE-constrained optimization using the conditional value-at-risk. SIAM J. Optim. **26**, 365–396 (2016)
19. Krokhmal, P., Palmquist, J., Uryasev, S.: Portfolio optimization with conditional value-at-risk objective and constraints. J. Risk **4**, 11–27 (2002)
20. Mafusalov, A., Uryasev, S.: CVaR (superquantile) norm: stochastic case. Eur. J. Oper. Res. **249**, 200–208 (2016)
21. Novak, E.: Quadrature formulas for monotone functions. Proc. Am. Math. Soc. **115**, 59–68 (1992)
22. Pòlya, G.: Anwendung des riemannschen integralbegriffes auf einig zahlentheoretische aufgaben. Arch. Math. Phys. **26**, 196–201 (1917)
23. Rabinowitz, P.: Gaussian integration in the presence of a singularity. SIAM J. Numer. Anal. **4**, 191–201 (1967)
24. Rockafellar, R.T., Royset, J.O.: Random variables, monotone relations, and convex analysis. Math. Program. **148**, 297–331 (2014)
25. Rockafellar, R.T., Royset, J.O.: Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. Eur. J. Oper. Res. **234**, 140–154 (2014)
26. Rockafellar, R.T., Royset, J.O.: Measures of residual risk with connections to regression, risk tracking, surrogate models and ambiguity. SIAM J. Optim. **25**, 1179–1208 (2015)
27. Rockafellar, R.T., Royset, J.O.: Superquantile/CVaR risk measures: second-order theory. Ann. Oper. Res. **262**, 3–28 (2018)
28. Rockafellar, R.T., Royset, J.O., Miranda, S.I.: Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. Eur. J. Oper. Res. **234**, 140–154 (2014)
29. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Finance **26**, 1443–1471 (2002)
30. Rockafellar, R.T., Uryasev, S.: The fundamental risk quadrangle in risk management, optimization and statistical estimation. Surv. Oper. Res. Manag. Sci. **18**, 33–53 (2013)
31. Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory, Second Edition, MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics, Philadelphia (2014)
32. Zavala, V.M.: Stochastic optimal control model for natural gas networks. Comput. Chem. Eng. **64**, 103–113 (2014)