SAND2022-1272J

Preprint

Sandia National Laboratories

National Nuclear Security Administration

Surrogate Modeling for Efficiently, Accurately and Conservatively Estimating Measures of Risk

J. D. Jakeman[1], D. P. Kouri[1], and J. G. Huerta[2]

[1]Optimization and Uncertainty Quantification, Sandia National Laboratories, Albuquerque, NM, 87123
[2]Statistical Sciences, Sandia National Laboratories, Albuquerque, NM, 87123

Abstract

We present a surrogate modeling framework for conservatively estimating measures of risk from limited realizations of an expensive physical experiment or computational simulation. Risk measures combine objective probabilities with the subjective values of a decision maker to quantify anticipated outcomes. Given a set of samples, we construct a surrogate model that produces estimates of risk measures that are always greater than their empirical approximations obtained from the training data. These surrogate models limit over-confidence in reliability and safety assessments and produce estimates of risk measures that converge much faster to the true value than purely sample-based estimates. We first detail the construction of conservative surrogate models that can be tailored to a stakeholder's risk preferences and then present an approach, based on stochastic orders, for constructing surrogate models that are conservative with respect to families of risk measures. Our surrogate models include biases that permit them to conservatively estimate the target risk measures. We provide theoretical results that show that these biases decay at the same rate as the L^2 error in the surrogate model. Numerical demonstrations confirm that risk-adapted surrogate models do indeed overestimate the target risk measures while converging at the expected rate.

Article Info

Correspondence:
J. D. Jakeman
jdjakem@sandia.gov

Keywords:
Uncertainty quantification, risk, average value-at-risk, stochastic dominance, decision making, surrogate models, reliability

1 Introduction

Model-based decision making is subject to various sources of uncertainty. Risk assessment is needed to quantify the effect of these uncertainties on the severity of predicted outcomes. Conceptually, *risk* refers to the possibility of events with undesirable consequences. Following [1], we define a risk as a triplet that includes a set of possible scenarios, the consequences or outcomes of those scenarios, and the probability of those outcomes. For example, consider the risk assessment of the truss structure depicted in Figure 1. Here, the set of possible scenarios are the anticipated loads $P_1, \ldots, P_6$ and the possible material properties A_1, A_2, E_1, E_2 of the truss. Each of these scenarios, which we lump into the vector X, lead to different outcomes that we characterize by the vertical displacement y at the mid-span. Consequently, the mid-span displacement is a random variable Y.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & 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.

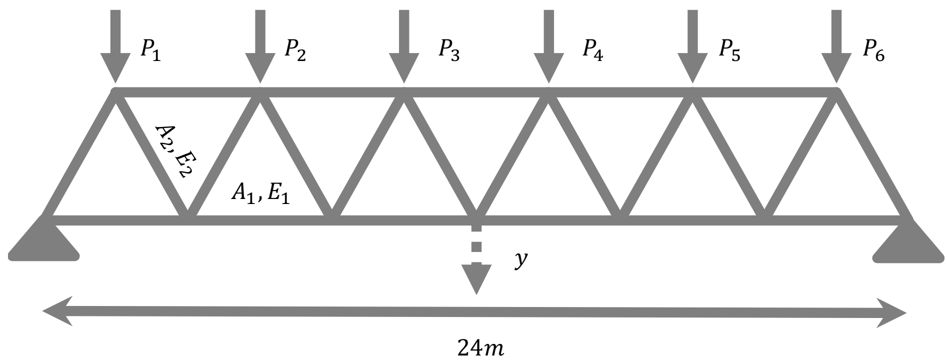

Figure 1: Two-dimensional truss structure with 23 bars and 13 nodes with uncertain loads $P_1, \ldots, P_6$ and material properties, specifically the cross sectional area A_i and Young's modulus E_i of the horizontal and diagonal bars, $i = 1, 2$, respectively. Our goal is to conservatively estimate a risk measure value associated with the vertical displacement y at the mid-span.

The above definition of risk is simply the statement of the likelihood and consequences of outcomes defined formally via the probability law of Y. However, rational decision making requires quantitative metrics, known as *risk measures*, that combine objective evidence with subjective values. For our truss example, assume that the decision maker associates vertical displacement with decreased safety and specifies a constant C as the maximum acceptable displacement. Risk measures provide a quantitative answer to the question "Is the output Y smaller than the threshold C?", when uncertainty may occasionally lead to Y exceeding C. If a risk measure is selected according to the preference of the stakeholder, then the value that it produces can be easily interpreted.

Numerous risk measures have been studied in the literature (cf. [2] and the references therein). For example, the expectation, $\mathbb{E}[Y]$, is neutral to risk and can be used to certify a truss when the decision maker does not place any value (or displeasure) on deviations from the expected behavior. In this case, a system is deemed safe if $\mathbb{E}[Y] \leq C$. The expectation is inappropriate for risk-averse stakeholders, as it does not capture any notion of variability and does not quantify rare events. The probability of failure (PoF), $\mathbb{P}(Y > C)$, is more applicable for measuring risk when the decision maker wants to avoid any failure, that is, any realization of Y above the predefined threshold C. However, when the stakeholder is interested in the magnitude of failure, a risk measure such as the average value-at-risk (AVaR)[1] may be more appropriate. See [3, 4, 5, 6, 7, 8] for recent works using risk measures to quantify and assess risk in engineering applications.

Estimating risk measures that characterize the nature of rare and undesirable events can be challenging. Monte Carlo (MC) sampling can be used to approximate risk measures. However, MC typically requires a large number of experiments or simulations, which can make accurately estimating the risk measure intractable when data generation is expensive. The cost of estimating tail statistics using MC can be reduced with importance sampling (IS) [9, 10, 11, 12] or subset simulation (SS) [13, 14, 15]. Unfortunately, similar to MC, the estimates of rare events using these methods converge slowly as the number of samples increases.

Replacing an expensive experiment or simulation with a inexpensive surrogate model, e.g. $\theta_0 + g(x, \theta)$, is a popular approach for reducing the cost of sample-based estimation. Given training data in the form of M input-output pairs (X_M, Y_M), the surrogate model parameters (θ_0, θ) are estimated so that

$$\theta_0 + g(X, \theta) = \widehat{Y} \approx Y = f(X).$$

Once constructed, the surrogate model can be sampled at negligible cost to evaluate measures of risk.

Various types of surrogate models can be used for estimating risk measures, including polynomials [16, 17, 18], Gaussian process or kriging models [19, 20, 21, 22, 23], neural networks [24, 25, 26], and reduced-order models [27, 28, 29]. However, these approaches introduce an error in the estimated risk measure value that must be quantified. In addition, surrogate models are often constructed using finitely many data points.

[1] Also called the expected shortfall, tail expectation, conditional value-at-risk, or superquantile.

Without imposing constraints on the estimation of (θ_0, θ), certain draws of training samples may lead to overestimating the risk measure value, while others can underestimate it.

Two general approaches are used to account for uncertainty in surrogate-based estimates of risk measures. The first approach generates confidence intervals (for example, obtained from the predictive variance of a Gaussian process) that can be used to inform decision making [30, 31, 32]. In this paper, we consider an alternative approach, which adds bias to the surrogate model so that it *conservatively* approximates a risk measure $\mathcal{R}(Y)$ in the sense that

$$\mathcal{R}(\widehat{Y}_M) \geq \mathcal{R}(Y_M), \tag{1}$$

where $\widehat{Y}_M$ denotes the values of the surrogate model obtained at the training samples X_M. Certain works, e.g. [33, 35], use heuristic strategies to add multiplicative and additive biases to surrogate models in an attempt to enforce (1). While others use elements of the risk quadrangle [37] to formulate a loss function that is guaranteed to train a surrogate model that will conservatively estimate a single risk measure [3, 8, 34, 36, 37].

In this paper, we develop a novel procedure to construct surrogate models that are guaranteed to conservatively estimate a family of risk measures. Such approaches are useful when it is not possible to elicit a single risk measure from the decision makers. This may arise when multiple stakeholders have competing risk preferences. To conservatively estimate multiple risk measures, we construct surrogate models using either first-order or second-order stochastic dominance constraints [38]. By enforcing the first-order stochastic dominance constraint

$$\mathbb{P}(\widehat{Y} > y) \geq \mathbb{P}(Y > y) \quad \forall\, y \in \mathbb{R},$$

we ensure that the surrogate model does not underestimate, in the sense of (1), a large class of risk measures including the PoF. Second-order dominance is a weaker condition that we use to effectively enforce the conservativeness of the surrogate model when first-order dominance is too risk averse.

The main contributions of this paper are:

- An overview of issues that must be considered when using surrogate models to estimate risk measures. Specific attention is given to procedures for eliciting risk measures that reflect the subjective risk-aversion of the application stakeholders and the need to tailor the surrogate models to evaluate their risk measures.

- The introduction of least-squares regression with stochastic dominance constraints to conservatively estimate a set of risk measures. This is useful when a single risk preference cannot be agreed upon. Numerical results also suggest that this approach is more likely to conservatively estimate a target risk measure, not just the empirical estimate from the training data.

- A rigorous theoretical and numerical analysis of the trade off between accuracy and bias for risk-adapted surrogate models. For example, we show that the error in the estimate of a target risk measure, obtained using polynomial chaos expansion (PCE) with stochastic dominance constraints, decays at the same rate as the root-mean-squared error for the PCE model. Similar novel results are also provided for existing approaches, based on the risk quadrangle, that target estimation of a single risk measure. These results demonstrate that increasing conservatism (bias) does not significantly decrease accuracy, but does reduce the likelihood of underestimation.

The subsequent sections are organized as follows. In Section 2, we discuss procedures for selecting risk measures based on the stakeholder's notion of regret or disutility and introduce the risk quadrangle and stochastic dominance. We then review the mathematical framework for using the risk quadrangle to construct conservative surrogate models tailored to a specific risk measure in Section 3. In Section 4, we present a stochastic dominance approach to construct surrogate models that are conservative with respect to broad classes of risk measures. Finally, we provide numerical examples that demonstrate the utility and efficacy of risk-adapted surrogate models in Section 5 and provide conclusions in Section 6. The proofs for all technical results are included in the appendix.

2 Risk Assessment for Decision Making

Risk assessment is typically undertaken to aid decision making. For example, risk assessment can be used to evaluate and compare the hazards associated with two different truss structures (Figure 1) with varying material properties. In this section, we discuss popular risk measures that quantify risk based on the subjective values of a decision maker. We then discuss procedures for selecting risk measures based on the stakeholder's notion of regret or disutility. To this end, we introduce the risk quadrangle, which codifies the connections between stochastic optimization and statistical estimation through the notions of risk, regret, deviation and error. Finally, we introduce stochastic dominance and discuss how it can be used when a single risk preference is difficult to elicit or when multiple stakeholders have competing subjective values. In the following, we denote the cumulative distribution function (CDF) of a random variable Y by $F_Y(t) := \mathbb{P}(Y \leq t)$.

2.1 Risk Measures

For our purpose, we define a risk measure as any real-valued function, $\mathcal{R}$, acting on a space of random variables, that satisfies $\mathcal{R}(C) = C$ for all constant values $C \in \mathbb{R}$. This requirement ensures that deterministic quantities are without risk. Practically speaking, risk measures provide a deterministic quantification of the distribution of possible outcomes associated with a risk and therefore are assumed to have the same units as the outcomes. There are numerous desirable axioms that a risk measure $\mathcal{R}$ should satisfy. For any two random variables Y and Y', five popular conditions from the literature are:

(R1) **Convexity:** $\mathcal{R}(tY + (1-t)Y') \leq t\mathcal{R}(Y) + (1-t)\mathcal{R}(Y')$ for all $t \in [0,1]$;

(R2) **Monotonicity:** If $Y \leq Y'$ almost surely, then $\mathcal{R}(Y) \leq \mathcal{R}(Y')$;

(R3) **Translation Equivariance:** $\mathcal{R}(Y + C) = \mathcal{R}(Y) + C$ for all $C \in \mathbb{R}$;

(R4) **Positive homogeneity:** $\mathcal{R}(tY) = t\mathcal{R}(Y)$ for all $t \geq 0$;

(R5) **Law Invariance:** If $F_Y(t) = F_{Y'}(t)$ for all $t \in \mathbb{R}$, then $\mathcal{R}(Y) = \mathcal{R}(Y')$.

A risk measure that satisfies (R1), (R2) and (R3) is called a convex risk measure [39]. If a convex risk measure additionally satisfies (R4), then it is said to be coherent [40]. The monotonicity condition (R2) ensures that any consequence Y that is always less than an alternative consequence Y' should have less risk when measured by $\mathcal{R}$, while the translation equivariance properties (R3) ensures that the deterministic shift C is without risk. Positive homogeneity (R4) ensures that if we change the units of Y, then the units of $\mathcal{R}(Y)$ also change in a consistent fashion. Finally, if $\mathcal{R}$ is law invariant (R5), then the risk measure value $\mathcal{R}(Y)$ is not affected by rearranging the scenarios of Y. Below, we list various popular law-invariant risk measures.

- **Expected Value**. $\mathcal{R}(Y) = \mathbb{E}[Y]$. This coherent risk measure is often referred to as risk neutral and should not be used in high-consequence applications when the extreme behavior of Y is important to quantify.

- **Mean-plus-standard-deviation**. $\mathcal{R}(Y) = \mathbb{E}[Y] + \lambda\mathbb{V}[Y]^{\frac{1}{2}}$, where $\lambda > 0$ is a user specified weight. This risk measure is often an improvement over mean-based risk as it quantifies the deviation of the outcomes from their average. However, it assigns the same weight to both positive and negative deviations, failing to account for skewness in the distribution. This risk measure does not satisfy the monotonicity condition (R2) and therefore is neither a coherent nor convex risk measure.

- **Worst-case**. $\mathcal{R}(Y) = \sup Y$. This coherent risk measure is extremely conservative and is typically difficult to estimate. It can often misinform decisions based on a single "bad" outcome because it does not consider how unlikely that outcome is or how "good" the alternative outcomes are.

- **Quantile**. $\mathcal{R}(Y) = q_p[Y] := \inf\{y \in \mathbb{R} \mid F_Y(y) \geq p\}$ for $p \in (0,1)$. The p^{th} quantile[2] provides a reasonable measure of conservativeness. In words, Y exceeds $q_p[Y]$ with probability $1-p$. Roughly

[2]Often called the value-at-risk (VaR) in financial applications.

speaking, when $p = 1$, $q_p[Y]$ is the largest value of Y and when $p = 0$, $q_p[Y]$ is the smallest value of Y. The quantile is closely related to the PoF and satisfies the equivalence: $q_p(Y) \le C \iff \mathbb{P}(X > C) \le 1 - p$. Unfortunately, the quantile does not characterize the weight of the distribution tail. Moreover, the quantile does not satisfy (R1) and therefore is neither a coherent nor convex risk measure.

- **Average value-at-risk**. $\mathcal{R}(Y) = \mathrm{AVaR}_p[Y] := \frac{1}{1-p}\int_p^1 q_\alpha[Y]\,\mathrm{d}\alpha$ for $p \in (0,1)$. AVaR is a coherent risk measure that quantifies the average of the tail in excess of $q_p[Y]$ and therefore provides a conservative alternative to the quantile (i.e., $\mathrm{AVaR}_p[Y] \ge q_p[Y]$ for all $p \in (0,1)$ and for all Y). When $p = 0$, AVaR returns the expected value and when $p \to 1$, AVaR returns the worst-case risk measure. Moreover, AVaR is closely related to the concept of buffered probability [41]. In particular, the buffered probability that Y exceeds C is defined as $\bar{p}_C(Y) = 1 - p$, where p is chosen so that $\mathrm{AVaR}_p[Y] = C$. Analogous to the relation between the quantile and the PoF, we have the equivalence: $\mathrm{AVaR}_p[Y] \le C \iff \bar{p}_C(Y) \le 1 - p$.

2.2 Eliciting Risk Preference

Risk measures should be tailored to the subjective beliefs of decision makers. We use an approach motivated by utility theory to mathematically model a decision maker's aversion to risk. A utility measure $\mathcal{U}$ is an increasing function that quantifies the decision maker's happiness associated with a risk. Similarly a disutility or regret measure $\mathcal{V}$ quantifies a decision makers displeasure with a risk and satisfies $\mathcal{V}(Y) = -\mathcal{U}(-Y)$. When large values of Y are undesirable, risk measures can be formulated intuitively from regret. Formally, a regret measure is a functional $\mathcal{V}$ that assigns each random outcome a scalar value that quantifies the displeasure for uncertainty associated with possible realizations of the outcomes Y. A common approach to defining $\mathcal{V}$ is with the expected regret $\mathcal{V}(Y) = \mathbb{E}[v(Y)]$ where $v : \mathbb{R} \to \mathbb{R}$ is an increasing function with $v(0) = 0$. Figure 2 demonstrates different types of regret functions v. In the subsequent discussion, we consider risk-averse regret measures that satisfy the basic mathematical properties: $\mathcal{V}$ is convex, lower semicontinuous and satisfies $\mathcal{V}(0) = 0$ and $\mathcal{V}(Y) > \mathbb{E}[Y]$ for all random variables $Y \neq 0$.

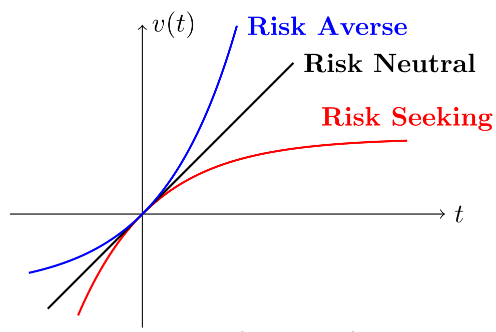

Figure 2: Regret models of risk preference. The decision maker is risk seeking when $v(t)$ is concave, risk averse when $v(t)$ is convex and risk neutral to risk when $v(t) = t$.

To construct a risk measure from $\mathcal{V}$, we employ *optimized certainty equivalent* (OCE) risk measures [42]. The OCE risk measures associated with $\mathcal{V}$ is defined by

$$\mathcal{R}(Y) = \inf_{d \in \mathbb{R}} \{d + \mathcal{V}(Y - d)\}.$$

To conceptualize the OCE risk measure, consider the truss depicted in Figure 1, where the regret $\mathcal{V}(Y)$ can be interpreted as the anticipated displacement beyond a reference value C. If we can preload the truss structure to achieve a certain mid-plane displacement d, then the anticipated displacement after preloading is $d + \mathcal{V}(Y - d)$. Consequently, $\mathcal{R}(Y)$ represents the anticipated displacement that would result from *optimal* preloading.

2.3 The Risk Quadrangle

In this section, we review how to generate risk-informed loss functions for statistical estimation using the *risk quadrangle* [37], depicted in Figure 3. The risk quadrangle components $\mathcal{R}$, $\mathcal{V}$, $\mathcal{E}$ and $\mathcal{D}$ are real-valued functions, acting on a space of random variables and are assumed to be lower semicontinuous and convex. Additional properties for these functionals are described in [37]. Three popular risk quadrangles are listed in Table 1 and Figure 4 depicts their associated regret functions.

$$\boxed{\begin{array}{lrcccll} & \text{Risk} & \mathcal{R} & \longleftrightarrow & \mathcal{D} & \text{Deviation} & \\ \text{Optimization} & & \uparrow\downarrow & \mathcal{S} & \downarrow\uparrow & & \text{Estimation} \\ & \text{Regret} & \mathcal{V} & \longleftrightarrow & \mathcal{E} & \text{Error} & \end{array}}$$

$$\mathcal{R}(Y) = \mathbb{E}[Y] + \mathcal{D}(Y) \qquad \mathcal{D}(Y) = \mathcal{R}(Y) - \mathbb{E}[Y]$$

$$\mathcal{V}(Y) = \mathbb{E}[Y] + \mathcal{E}(Y) \qquad \mathcal{E}(Y) = \mathcal{V}(Y) - \mathbb{E}[Y]$$

$$\mathcal{R}(Y) = \inf_{t\in\mathbb{R}}\{t + \mathcal{V}(Y-t)\} \qquad \mathcal{D}(Y) = \inf_{t\in\mathbb{R}} \mathcal{E}(Y-t)$$

$$\mathcal{S}(Y) = \arg\min_{t\in\mathbb{R}}\{t + \mathcal{V}(Y-t)\} = \arg\min_{t\in\mathbb{R}} \mathcal{E}(Y-t)$$

Figure 3: The risk quadrangle.

Table 1: Examples of different risk quadrangles

Safety Margins Risk Quadrangle	**p-th Quantile Risk Quadrangle**	**Entropic Risk Quadrangle**
$\mathcal{V}(Y) = \mathbb{E}[Y] + \lambda\|Y\|_2$	$\mathcal{V}(Y) = \frac{1}{1-p}\mathbb{E}[\max\{0, Y\}]$	$\mathcal{V}(Y) = \mathbb{E}[\exp(Y) - 1]$
$\mathcal{R}(Y) = \mathbb{E}[Y] + \lambda\mathbb{V}[Y]^{\frac{1}{2}}$	$\mathcal{R}(Y) = \mathrm{AVaR}_p\,[Y]$	$\mathcal{R}(Y) = \log\mathbb{E}[\exp(Y)]$
$\mathcal{E}(Y) = \lambda\|Y\|_2$	$\mathcal{E}(Y) = \mathbb{E}[\frac{p}{1-p}\max\{0, Y\} + \max\{0, -Y\}]$	$\mathcal{E}(Y) = \mathbb{E}[\exp(Y) - Y - 1]$
$\mathcal{D}(Y) = \lambda\mathbb{V}[Y]^{\frac{1}{2}}$	$\mathcal{D}(Y) = \mathrm{AVaR}_p\,[Y - \mathbb{E}[Y]]$	$\mathcal{D}(Y) = \log\mathbb{E}[\exp(Y - \mathbb{E}[Y])]$
$\mathcal{S}(Y) = \mathbb{E}[Y]$	$\mathcal{S}(Y) = q_p[Y]$	$\mathcal{S}(Y) = \log\mathbb{E}[Y]$

Once determined the regret measure can be used to uniquely define an OCE risk measure, as discussed in the previous section and shown in Figure 3. The risk quadrangle also provides additional connections between measures of regret, deviation, and error. Specifically, the regret measure uniquely determines an error measure $\mathcal{E}$ that can be used as a loss function to train surrogate models tailored to estimating $\mathcal{R}$. The error measure quantifies the proximity of a random quantity to zero and satisfy $\mathcal{E}(Y) = 0$ if and only if $Y = 0$. The error measure also defines a deviation measure $\mathcal{D}$, which is a (possibly non-symmetric) generalization of the standard deviation that quantifies the variability of the random quantity and satisfies $\mathcal{D}(C) = 0$ if and only if C is constant. Finally, the optimal values t of the optimization problems that define $\mathcal{R}$ and $\mathcal{D}$ are encapsulated in the set-valued map $\mathcal{S}$, which is called the statistic.[3]

2.4 Stochastic Dominance

The choice of risk measure should be tailored to the stakeholder's risk preference. Unfortunately, eliciting a single risk preference can be difficult. In such situations, it may be more appropriate to assess risk with respect to a class of risk measures or even entire distributions. Stochastic dominance can be used in such cases. For this work, we restrict our attention to dominance with respect to the first and second stochastic orders.

The random variable Y is said to dominate another random variable Y' with respect to the first stochastic order, denoted $Y \succeq_{(1)} Y'$, if

$$F_Y(t) \leq F_{Y'}(t) \quad \forall\, t \in \mathbb{R}. \tag{2}$$

The first-order stochastic dominance (FSD) condition (2) is equivalent to

$$\mathbb{E}[u(Y)] \leq \mathbb{E}[u(Y')]$$

[3] It can be difficult to produce meaningful regret and error measures from a prescribed risk or deviation measure. Fortunately, it is often more intuitive to elicit a regret or utility measure and use it to define the other risk quadrangle quantities through the relationships in Figure 3.

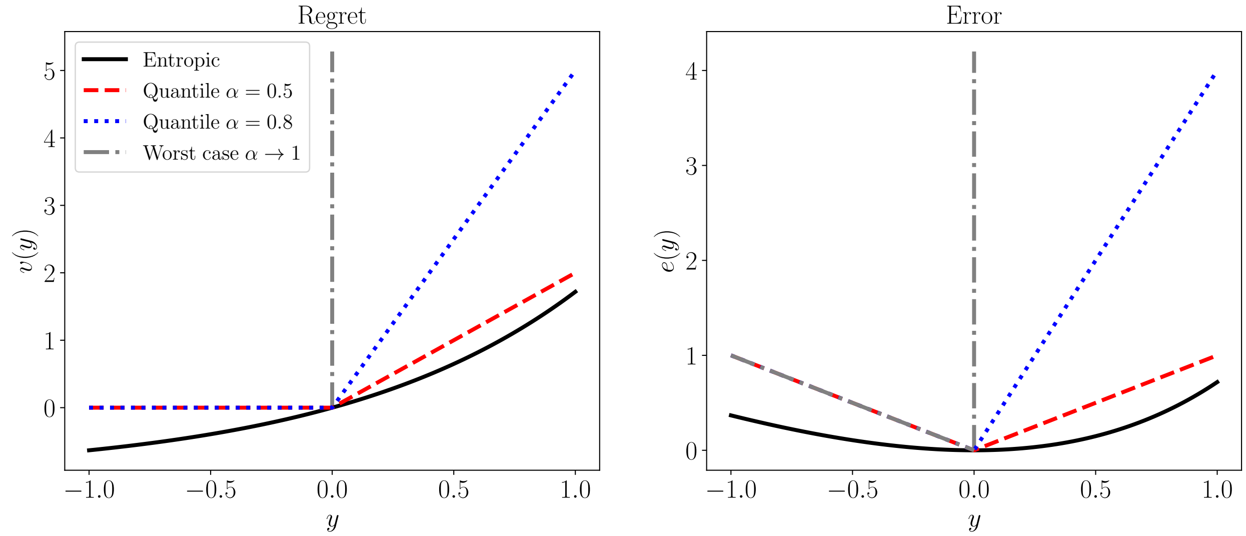

Figure 4: (left) Popular regret functions $v(y)$. (Right) Associated error functions $e(y)$ codified by the risk quadrangle.

for all nondecreasing functions $u : \mathbb{R} \to \mathbb{R}$ for which the expectations exist. On the other hand, Y dominates Y' with respect to the second stochastic order, denoted $Y \succeq_{(2)} Y'$, if

$$\int_{-\infty}^{t} F_Y(\eta)\,\mathrm{d}\eta \le \int_{-\infty}^{t} F_{Y'}(\eta)\,\mathrm{d}\eta \quad \forall\, t \in \mathbb{R}$$
$$\iff \quad \mathbb{E}[\max\{0, t - Y\}] \le \mathbb{E}[\max\{0, t - Y'\}] \quad \forall\, t \in \mathbb{R}.$$

Second-order stochastic dominance (SSD) focuses on properties of the left tail, i.e., small values of Y. For engineering and science applications, we are typically interested in undesirably large values of Y. For this reason, we employ the so-called increasing convex order defined by

$$Y' \succeq_{\mathrm{icx}} Y \quad \iff \quad \mathbb{E}[u(Y)] \le \mathbb{E}[u(Y')] \tag{3}$$

for all nondecreasing convex functions $u : \mathbb{R} \to \mathbb{R}$ for which the expectations exist. In words, if $Y' \succeq_{\mathrm{icx}} Y$, then the expected utility of Y is always smaller than the expected utility of Y'. The increasing convex order is related to SSD as follows

$$Y' \succeq_{\mathrm{icx}} Y \quad \iff \quad -Y \succeq_{(2)} -Y'.$$

In particular, we have that $Y' \succeq_{\mathrm{icx}} Y$ if and only if

$$\mathbb{E}[\max\{0, Y - t\}] \le \mathbb{E}[\max\{0, Y' - t\}] \quad \forall\, t \in \mathbb{R}.$$

As such, we will refer to dominance with respect to the increasing convex order simply as SSD. It is clear from the utility representation of FSD and SSD that if $Y \succeq_{(1)} Y'$, then $-Y \succeq_{(2)} -Y'$. However, the converse is not true in general.

Figure 5 illustrates the difference between FSD and SSD. The FSD relation $Y_2 \succeq_{(1)} Y_1$ ensures that the CDF of Y_2 lies to the right of the CDF Y_1 in the left plot. In particular, $Y_2 \succeq_{(1)} Y_1$ implies that the PoF of Y_2 will be greater than that of Y_1 for any threshold C. We will use this property to construct surrogate models that conservatively estimate the PoF in Section 4.

Despite being less risk averse than FSD, the SSD relationship $-Y_2 \succeq_{(2)} -Y_3$ is more risk-averse than requiring

$$\mathcal{R}(Y_3) \ge \mathcal{R}(Y_2) \tag{4}$$

for many single risk measures $\mathcal{R}$. In fact, the SSD relation $-Y_2 \succeq_{(2)} -Y_3$ ensures that $\mathrm{AVaR}_p[Y_3]$ is larger than $\mathrm{AVaR}_p[Y_2]$ for any confidence level $p \in (0, 1)$. Consequently, if the underlying probability space is

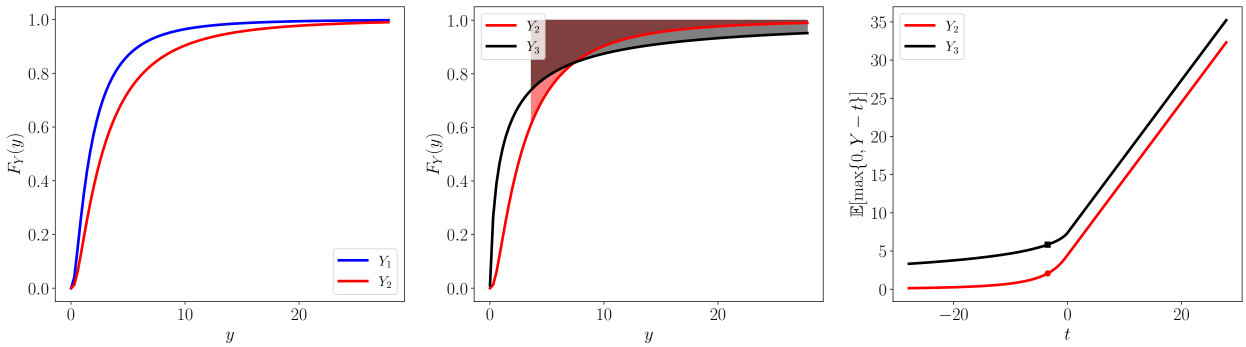

Figure 5: Comparison of first- and second-order stochastic dominance. (Left) $Y_2 \succeq_{(1)} Y_1$. (Middle) $-Y_2 \succeq_{(2)} -Y_3$, but there is no first-order dominance. (Right) The values $\mathbb{E}[\max\{0, Y - t\}]$ confirming second-order dominance. The average of the shaded regions in the middle plot equals the value of the conditional expectation for Y_2 and Y_3 in excess of $t = 3.5$, which are shown as dots on the right plot.

nonatomic, then (4) holds for all risk measures that are proper, lower semicontinuous and satisfy (R1), (R2), (R3), and (R5) [2, Th. 6.51]. We will use SSD to construct surrogate models that conservatively estimate the AVaR and other risk measures in Section 4.

3 Risk-Adapted Surrogate Modeling

The risk quadrangle provides a rigorous connection between stochastic optimization and statistical estimation that facilitates the translation of a stakeholder's preferences to mathematical definitions of risk, deviation, error and regret. In this section, we leverage these connections to develop a strategy for constructing surrogate models that accurately and conservatively estimate a chosen risk measure using data obtained from models of the form

$$Y = f(X),$$

where $f(\cdot)$ represents a (potentially unknown) consequence or loss associated with input or environmental variables $X = (X_1, \ldots, X_D) \in \Omega_X \subseteq \mathbb{R}^D$ and $Y \in \Omega_Y \subseteq \mathbb{R}$ is the observed consequence. We conclude this section with the derivation of an error bound for the surrogate-based risk estimate and with detailed numerical procedures for constructing these surrogate models.

3.1 Constructing Risk-Adapted Surrogate Models

The risk quadrangle can be used to formulate regression problems tailored to the regret measure elicited from the decision maker by exploiting the connections between regret and error shown in Figure 3 [36, 34]. Specifically, error measures use the stakeholder's notion of regret to measure how non-zero a random variable is through the relationship $\mathcal{E}(Y) = \mathcal{V}(Y) - \mathbb{E}[Y]$. The right panel of Figure 4 depicts the error functions $e(y) = v(y) - y$ associated with the regret functions in the left plot of the same figure. It is clear that different risk preferences lead to different error measures and thus penalize the residuals $Y - \theta_0 - g(X, \theta)$ differently, according to magnitude and sign. The worst-case regret is associated with an error function that assigns an infinitely large penalty to any positive value. Noting that we define residuals as the signed difference $Y - \theta_0 - g(X, \theta)$, we see that the worst-case error function attempts to produce a surrogate model that is always larger than the data. The other error functions depicted also assign larger penalties to positive values, relative to negative values, but to differing degrees.

Given an error measure that was generated from a regret measure, we can calibrate the surrogate model by a two-step procedure.

1. Compute θ_0' and $\theta^\star \in \Theta$ by solving

$$\min_{\theta_0 \in \mathbb{R},\, \theta \in \Theta} \mathcal{E}(Y - \theta_0 - g(X, \theta)). \tag{5}$$

2. Replace θ_0' with

$$\theta_0^\star := \mathcal{R}(Y - g(X, \theta^\star)). \tag{6}$$

Using this two-step procedure is guaranteed to produce a surrogate that satisfies

$$\mathcal{R}(Y) \leq \theta_0^\star + \mathcal{R}(g(X, \theta^\star)). \tag{7}$$

See Appendix A for a detailed proof.

To illustrate this procedure consider the safety margins risk quadrangle in Table 1 as an example. For the risk quadrangle the mean-plus-standard-deviation risk measure is estimated using least-squares regression, i.e.,

$$\min_{\theta_0 \in \mathbb{R},\, \theta \in \Theta} \sum_{m=1}^{M} \pi^{(m)} \left[y^{(m)} - \theta_0 - g(x^{(m)}, \theta) \right]^2. \tag{8}$$

The shift $\theta_0^\star$ is then compute using the residuals

$$R = [r^{(1)}, \ldots, r^{(M)}] \qquad\qquad r^{(m)} := y^{(m)} - g(x^{(m)}, \theta^\star) \quad \text{for} \quad m = 1, \ldots, M$$

and setting

$$\theta_0^\star = \sum_{m=1}^{M} \pi^{(m)} r^{(m)} + \lambda \left[\sum_{m=1}^{M} \pi^{(m)} \left(r^{(m)} - \sum_{m=1}^{M} \pi^{(m)} r^{(m)} \right)^2 \right]^{\frac{1}{2}}.$$

Similarly for the quantile risk quadrangle, we solve the quantile regression problem

$$\min_{\theta_0 \in \mathbb{R},\, \theta \in \Theta_k} \sum_{m=1}^{M} \pi^{(m)} \gamma_p \left[y^{(m)} - \theta_0 - g(x^{(m)}, \theta) \right]$$
$$\text{where} \quad \gamma_p[u] = p \max(0, u) + (1-p) \max(0, -u).$$

and set

$$\theta_0^\star = q_p(R) + \frac{1}{1-p} \sum_{m=k+1}^{M} \pi^{(m)} (r^{(m)} - q_p(R)),$$

where the residuals R are sorted from smallest to largest so that the p-th quantile is $q_p(R) = r^{(k)}$ such that k is the smallest index satisfying $\pi^{(1)} + \ldots + \pi^{(k)} \geq 1 - p$.

Note that when g is linear in θ and $\Theta = \mathbb{R}^K$, the quantile regression problem can be efficiently solved as a linear program [49]. More complicated, nondifferentiable optimization techniques are required when g depends nonlinearly on θ [50].

3.2 Orthonormal Surrogate Models Using the Risk Quadrangle

The risk quadrangle can be applied to most popular approximation strategies including polynomial chaos expansions (PCE) [43, 44], low-rank tensor decompositions [45, 46], and neural networks [47, 48]. In this section, we derive error bounds for estimates of $\mathcal{R}(Y)$ using surrogate models based on orthonormal basis functions, such as PCEs. For this, we assume that the probability laws of X and Y are non-atomic. We investigate surrogate models with the form

$$f(x) \approx \theta_0 + \sum_{k=1}^{K} \psi_k(x) \theta_k := \theta_0 + g_K(x, \theta), \tag{9}$$

where K is the truncation order, $\Theta = \Theta_K := \mathbb{R}^K$, $\theta = [\theta_1, \ldots, \theta_K]^\top$, and $\{\psi_k\}_{k=1}^\infty$ is an orthonormal system. In particular, $\{\psi_k\}_{k=1}^\infty$ satisfies $\mathbb{E}[\psi_k(X)] = 0$ and $\mathbb{E}[\psi_k(X)\psi_\ell(X)] = \delta_{k\ell}$ and the Fourier coefficients associated with this basis are $\overline{\theta}_0 := \mathbb{E}[Y]$ and $\overline{\theta}_k := \mathbb{E}[\psi_k Y]$ for $k = 1, 2, \ldots, K$. Provided Y has finite variance, the surrogate model (9) with Fourier coefficients $(\theta_0, \theta) = (\overline{\theta}_0, \overline{\theta}_{1:K})$, where $\overline{\theta}_{1:K} = [\overline{\theta}_1, \ldots, \overline{\theta}_K]^\top$, satisfies

$$\mathbb{E}[|f(X) - \overline{\theta}_0 - g_K(X, \overline{\theta}_{1:K})|^2] \to 0 \qquad \text{as} \qquad K \to \infty.$$

In the following result, we show that the error for the estimate of $\mathcal{R}(Y)$, obtained using g_K, decays at a similar rate as the truncation error of g_K, i.e.,

$$\mathbb{E}[(Y - \overline{\theta}_0 - g_K(X, \overline{\theta}))^2] = \tau_K := \sum_{k=K+1}^{\infty} \overline{\theta}_k^2,$$

where τ_K is the tail energy associated with the order K approximation. To state our next result, we make the following observation. If $\mathcal{R}$ is positively homogeneous, $\theta \in \Theta_K$ and $\theta_0 = \mathcal{R}(Y - g_K(X, \theta))$, then the following bounds hold

$$\begin{aligned} 0 &\le \theta_0 + \mathcal{R}(g_K(X, \theta)) - \mathcal{R}(Y) \\ &= \mathcal{D}(Y - g_K(X, \theta)) + \mathcal{D}(g_K(X, \theta)) - \mathcal{D}(Y) \\ &\le \mathcal{D}(Y - g_K(X, \theta)) + \mathcal{D}(g_K(X, \theta) - Y). \end{aligned} \tag{10}$$

This follows from the convexity and positive homogeneity of $\mathcal{R}$ and $\mathcal{D}$. In particular, for any two random variables, Y and Y', we have that

$$\mathcal{R}(Y) = \mathcal{R}(Y' + (Y - Y')) \le \mathcal{R}(Y') + \mathcal{R}(Y - Y') \tag{11}$$

and similarly for $\mathcal{D}$. Using (10), we can relate our risk estimation error with the mean-squared error.

Proposition 3.1. *Suppose the error measure $\mathcal{E}$ is positively homogeneous and that there exists $C > 0$ and $\alpha > 0$ such that*

$$\mathcal{E}(X) \le C\mathbb{E}[X^2]^\alpha. \tag{12}$$

Moreover, let $\theta^* \in \Theta_K$ solve the modified deviation minimization problem

$$\min_{\theta \in \Theta_K} \{\mathcal{D}(Y - g_K(X, \theta)) + \mathcal{D}(g_K(X, \theta))\} \tag{13}$$

and let $\theta_0^\star = \mathcal{R}(Y - g_K(X, \theta^\star))$. Then, the surrogate model $\theta_0^\star + g_K(X, \theta^\star)$ satisfies the following bound

$$0 \le \theta_0^\star + \mathcal{R}(g_K(X, \theta^\star)) - \mathcal{R}(Y) \le 2C\tau_K^\alpha.$$

The modified deviation minimization problem (13) includes the term $\mathcal{D}(g_K(X, \theta))$, which can be interpreted as a regularizer that balances the residual deviation and the model variability. Moreover, the problem (13) directly minimizes the difference between the surrogate and true risk values as demonstrated by (10). Finally, when $\mathcal{E}$ is positively homogeneous, Proposition 3.1 ensure that the error in the risk evaluation computed by solving (13) decays to zero as K increases. In particular, we recover the spectral convergence of the least-squares estimator. However, the requirement that $\mathcal{E}$ satisfies (12) may not hold in general.

Note that for the safety margins and quantile risk quadrangles (see Table 1) the assumptions of Proposition 3.1 hold with $\alpha = \frac{1}{2}$ and $C = 1$ for the mean-plus-standard-deviation, and $\alpha = \frac{1}{2}$ and $C = \frac{p}{1-p}$ for AVaR, respectively.

4 Risk-Averse Surrogates Using Stochastic Dominance

When the decision maker's risk preference is difficult to quantify with a single regret measure, it may be more appropriate to construct approximations that enforce conservativeness with respect to a class of risk measures or even entire distributions. Stochastic orders can be used to achieve this goal. For this discussion, we restrict our attention to the first and second stochastic orders. However, one could consider higher stochastic orders, or more generally integral stochastic orders [2, Ch. 6.3.7], within the same framework.

With the goal of constructing accurate surrogate models that conservatively estimate the empirical CDF generated by the training data, we consider the estimation problem

$$\min_{\theta_0\in\mathbb{R},\,\theta\in\Theta} \ \frac{1}{2}\mathbb{E}[|Y-\theta_0-g(X,\theta)|^2] \tag{14a}$$

$$\text{subject to} \quad \theta_0+g(X,\theta)\succeq_{(1)} Y. \tag{14b}$$

Given $\theta_0^\star\in\mathbb{R}$ and $\theta^\star\in\Theta$ that solve (14), we have the ordering $\mathbb{P}(\theta_0^\star+g(X,\theta^\star)>t)\geq\mathbb{P}(Y>t) \quad \forall\, t\in\mathbb{R}$ and therefore $\theta_0^\star+g(X,\theta^\star)$ will produce conservative estimates of the PoF of Y for all threshold C. In addition, Theorem 6.50 in [2] ensures that $\mathcal{R}(\theta_0^\star+g(X,\theta^\star))\geq\mathcal{R}(Y)$ for all law-invariant (R5) and monotonic (R2) risk measures.

In a similar fashion, we can construct surrogate models that are second-order dominant by solving the estimation problem

$$\min_{\theta_0\in\mathbb{R},\,\theta\in\Theta} \ \frac{1}{2}\mathbb{E}[|Y-\theta_0-g(X,\theta)|^2] \tag{15a}$$

$$\text{subject to} \quad -Y\succeq_{(2)} -(\theta_0+g(X,\theta)). \tag{15b}$$

See [38] for more information on stochastic-order constrained optimization problems. As discussed in Section 2.4, the SSD constraint (15b) ensures that the surrogate model overestimates Y with respect to all risk measures that satisfy (R1), (R2), (R3), and (R5). Note that if Y is exactly represented by the model $\theta_0+g(X,\theta)$, then both (14) and (15) will produce parameter estimates that exactly reproduce Y. That is, the least-squares loss function will be zero and the constraints will be satisfied with equality.

4.1 Orthonormal Surrogate Models Using Stochastic-Order Constraints

In this section, we determine an upper bound on the risk estimation error obtained for the truncated orthonormal surrogate model g_K in Section 3.2, calibrated using stochastic dominance constrained least-squares estimation. The following proposition (proved in Appendix B) provides a bound on the surrogate model error when the stochastic dominance constraints are enforced on a compact subset of $\mathbb{R}$. In particular, given a compact set $S\subset\mathbb{R}$, we enforce the FSD constraints

$$\mathbb{P}(\theta_0+g(X,\theta)>t)\geq\mathbb{P}(Y>t) \quad \forall\, t\in S \tag{16}$$

and the SSD constraints

$$\mathbb{E}[\max\{0,\theta_0+g(X,\theta)-t\}]\geq\mathbb{E}[\max\{0,Y-t\}] \quad \forall\, t\in S. \tag{17}$$

Proposition 4.1. *Let $S\subset\mathbb{R}$ be a compact set and $f:\mathbb{R}^D\to\mathbb{R}$ be a continuous function. Moreover, suppose that the unconstrained least-squares approximation $\bar\theta_0+g_K(\cdot,\bar\theta)$ satisfies the error bound*

$$\sup_{x\in f^{-1}(S)} |f(x)-(\bar\theta_0+g_K(x,\bar\theta))|\leq\delta_K \tag{18}$$

with $\delta_K\downarrow 0$ as $K\to\infty$. Here, $f^{-1}(S):=\{x\in\mathbb{R}^D \mid f(x)\in S\}$. Then, the solution, $(\theta_0^\star,\theta^\star)\in\mathbb{R}\times\Theta_K$, to the modified stochastic dominance constrained least-squares problem, with the FSD (or SSD) constraint replaced by (16) *(or* (17)*) satisfies*

$$\mathbb{E}[(Y-\theta_0^\star-g_K(X,\theta^\star))^2]\leq\tau_K+\delta_K^2$$

and

$$\|(\theta_0^\star,\theta^\star)-(\bar\theta_0,\bar\theta)\|_2\leq\delta_K.$$

Here, $\|\cdot\|_2$ denotes the Euclidean norm.

Proposition 4.1 demonstrates that the approximation error using FSD and SSD constraints is no worse than the least-squares error plus a metric of the uniform approximation quality of the least-squares model given by (18). See Chapter 3 in [51] for a discussion of best uniform approximations and the applicability of the condition (18).

4.2 Constructing Surrogate Models that Satisfy Stochastic Order Constraints

In this section, we discuss the optimization problems that must be solved to construct surrogate models that satisfy stochastic-order constraints. Given access to a finite set of data (X_M, Y_M) of size M, we replace the FSD and SSD constraints in (14) and (15) by

$$\sum_{m=1}^{M} \pi^{(m)} \mathbb{1}_{(-\infty,0]}(g(x^{(m)},\theta) - g(x^{(i)},\theta)) \leq \sum_{m=1}^{M} \pi^{(m)} \mathbb{1}_{(-\infty,0]}(y^{(m)} - \theta_0 - g(x^{(i)},\theta)) \tag{19}$$

and

$$\sum_{m=1}^{M} \pi^{(m)} \max\{0, g(x^{(m)},\theta) - g(x^{(i)},\theta)\} \leq \sum_{m=1}^{M} \pi^{(m)} \max\{0, y^{(m)} - \theta_0 - g(x^{(i)},\theta)\}$$

for $i = 1, \ldots, M$, respectively. Note that since $y^{(m)}$ and $g(x^{(m)},\theta)$ only assume a discrete set of values, it is sufficient to enforce the dominance constraints on the discrete set of values $t \in \{\theta_0 + g(x^{(m)},\theta)\}_{m=1}^{M}$.

4.2.1 First-Order Stochastic Dominance

In practice, solving the sampled FSD-constrained least-squares problem (14) is challenging due to the discontinuity introduced by the Heaviside function $\mathbb{1}_{(-\infty,0]}$ in (19). One approach to solving (14) is to reformulate the problem as a mixed integer optimization problem. However, if g depends nonlinearly on θ, then the resulting problem would be extremely challenging to solve due to problem size and potential nonconvexity [52]. To overcome this complication, we smooth the discontinuous constraints similar to [53]. In particular, we replace $\mathbb{1}_{(-\infty,0]}$ with two smooth approximations h_1, $h_2 : \mathbb{R} \to [0,1]$ and solve

$$\min_{\theta_0\in\mathbb{R},\, \theta\in\Theta} \ \frac{1}{2} \sum_{m=1}^{M} \pi^{(m)} (y^{(m)} - \theta_0 - g(x^{(m)},\theta))^2 \tag{20a}$$

subject to

$$\begin{aligned} &\sum_{m=1}^{M} \pi^{(m)} h_1(g(x^{(m)},\theta) - g(x^{(i)},\theta)) \leq \sum_{m=1}^{M} \pi^{(m)} h_2(y^{(m)} - \theta_0 - g(x^{(i)},\theta)) \\ &i = 1, \ldots, M. \end{aligned} \tag{20b}$$

If h_1 and h_2 satisfy

$$\mathbb{1}_{(-\infty,0]}(t) \leq h_1(t) \qquad \text{and} \qquad h_2(t) \leq \mathbb{1}_{(-\infty,0]}(t) \quad \forall\, t \in \mathbb{R}$$

and if $\theta_0^\star \in \mathbb{R}$ and $\theta^\star \in \Theta$ solve (20), then the original constraints (19) are also satisfied. Consequently, the resulting surrogate model will be conservative with respect to the data $\{y^{(m)}\}$. Unfortunately, such h_1 and h_2 often lead to an inconsistent estimation problem. For example, suppose Y is exactly represented by the model $\theta_0 + g(X,\theta)$, then the smoothed constraint in (20) may not be satisfied at the optimal values of θ_0 and θ, i.e., the feasible set for (20) is smaller than the feasible set defined by (19).

4.2.2 Second-Order Stochastic Dominance

To approximately solve the sampled SSD-constrained least-squares problem (15), one can reformulate the problem by adding slack variables. This approach increases the number of constraints from M to $O(M^2)$, making the resulting problem extremely difficult to solve when M is large. Similar to the FSD approach, we instead approximate a solution to (15) by employing a smooth approximation d to the positive-part function

$x \mapsto \max\{0, x\}$. In particular, we solve the optimization problem

$$\min_{\theta_0 \in \mathbb{R},\, \theta \in \Theta} \frac{1}{2} \sum_{m=1}^{M} \pi^{(m)} (y^{(m)} - \theta_0 - g(x^{(m)}, \theta))^2 \tag{21a}$$

subject to

$$\sum_{m=1}^{M} \pi^{(m)} d(g(x^{(m)}, \theta) - g(x^{(i)}, \theta)) \geq \sum_{m=1}^{M} \pi^{(m)} d(y^{(m)} - \theta_0 - g(x^{(i)}, \theta)) \quad i = 1, \ldots, M. \tag{21b}$$

See [38] for a discussion of sample-based approximation for optimization problems with SSD constraints. We note that (21) differs from the problems considered in [38] due to the presence of the minus sign in (15b). Additionally, we note that (21) is not convex as a result of the constraint (21b), even if the model g is linear in θ.

5 Numerical Examples

To confirm our theoretical results, we now present multiple numerical examples that highlight the properties and utility of risk-adapted surrogate models. All numerical results were generated using PyApprox [54] which is a python package for approximation and probabilitic analysis of data. In all examples, unless otherwise stated, we construct total-degree polynomial chaos expansions consisting of tensor product basis functions. Specifically, for each D-dimensional random vector X, we construct a univariate polynomial basis $\{\psi_{k,d}\}$, which is orthonormal with respect to the PDF of the d-th variable and set

$$\psi_\lambda(x) = \prod_{d=1}^{D} \psi_{\lambda_d, d}(x_d),$$

where $\lambda = [\lambda_1, \ldots, \lambda_D]$ is a multi-index specifying the degree of the basis in each dimension. A total-degree polynomial expansion of degree Q is then given by

$$\theta_0 + \sum_{\lambda \in \Lambda} \psi_\lambda(x) \theta_\lambda \qquad \text{with} \qquad \Lambda := \left\{ \lambda \mid 1 \leq \sum_{d=1}^{D} \lambda_d \leq Q \right\}, \tag{22}$$

where there is a one-to-one correspondence between the basis functions $\{\psi_\lambda\}_{\lambda \in \Lambda}$ and the basis functions $\{\psi_k\}_{k=1}^K$ used in (9). Here, the number of basis functions K is given by $K = |\Lambda| = \binom{D+Q}{D} - 1$. For canonical random variables (e.g., normal or uniform), analytical expressions exist for the recursion coefficients needed to construct the univariate orthonormal polynomials. For other variables (e.g. Gumbel), recursion coefficients must be computed numerically [55].

For our FSD results, we set $h_1 = h_2 = h$, which can theoretically produce a less conservative surrogate model. However, we found that the impact on the optimal solution is negligible. We choose the smoothed Heaviside function h based on the first derivative of the smoothed positive part function

$$d(x; \delta, \epsilon) = x + \epsilon \log\left(-\frac{x + \delta}{\epsilon}\right), \tag{23}$$

which we use for the SSD results. The corresponding approximate right Heaviside function used to produce the FSD results is given by

$$h_r(x; \delta_r, \epsilon) = \frac{\partial}{\partial x} d(x; \delta_r, \epsilon) = \frac{1}{1 + \exp\left(-\frac{x+\delta_r}{\epsilon}\right)}. \tag{24}$$

Here, δ_r and ϵ are tunable parameters. As $\epsilon \to 0$ the right Heaviside function approaches $\mathbb{1}_{[0,\infty)}(\cdot)$, so we use the left Heaviside function $h(x) = h_l(x; \delta_l, \epsilon) = h_r(-x; -\delta_l, \epsilon)$. When $\delta_l = 0$, we have $h(0) = \frac{1}{2}$ for all ϵ, which can lead to under-prediction of rare events. We found that setting $\delta_l = -\epsilon$, which results in $h(0) = \frac{1}{1+e^{-1}} \approx 0.73$ produced more conservative models. Unless otherwise stated, we set $\epsilon = 10^{-2}$. Note that our smoothed Heaviside function is never equal to the true Heaviside function when ϵ is non-zero.

We investigated the performance of alternative smooth approximations that are only inaccurate on a local interval of size ϵ. However, we found that solving (20) using these smoothers was very challenging. The Jacobian of the constraints are often low-rank because the gradient of the local Heaviside approximation is zero outside of an ϵ-sized interval. This prevents the optimization algorithm from choosing steps that lead to solutions that satisfy the constraints. Although the smooth Heaviside function that we use performed better than the alternatives that we investigated, the resulting optimization problem is still challenging to solve especially for small ϵ. Future work is needed to develop a robust formulation for enforcing FSD and SSD constraints.

All FSD and SSD surrogate models reported below are obtained by solving the associated regression problems using the SciPy [57] implementation of the interior point trust-region algorithm from [58]. Both regression problems are nonconvex and thus we are only guaranteed to find a local minima that depends on the initial guess passed to the optimizer. We use a three step procedure to determine a reasonable feasible initial guess. The first step computes the unconstrained least-squares (LstSq) solution. The second step then computes the magnitude of the largest residual between the least-squares model and the training data. Finally, the third step adds the size of the largest residual to the coefficient corresponding to the constant basis term of the PCE so that the initial PCE always lies above the training data.

5.1 Risk Estimation for a Log-Normal Output

Let X be a D-dimensional normal random vector with mean μ and covariance Σ. We are interested in computing risk measures from outcomes generated by the model

$$Y = f(X) = \exp(a^\top X), \tag{25}$$

where $a = [a_1, \dots, a_D]^\top$ and $a_d > 0$ for $d = 1, \dots, D$. In this setting, the random outcome Y is a log-normal random variable for which we can compute various risk measures and dominance constraints analytically (see, e.g., [56, Sect. 3.2]). We construct Hermite-PCE surrogate models using the training samples $\{x^{(m)}\}_{m=1}^M$ that we drew randomly using MC. We then compute the PCE-based estimates of risk measures using 10^5 Quasi Monte Carlo (QMC) evaluations of the surrogate model. We employ inverse transform sampling to ensure that the QMC samples integrate with respect to the distribution of Y.

To clarify the forthcoming discussion, we denote the exact output Y by Y_∞, and recall that Y_M denotes the M-sample empirical approximation of Y_∞. We further recall that $Y_{M,N}$ denotes the N-sample empirical approximation of the surrogate data $\widehat{Y}$, trained using M samples. As a special case of $Y_{M,N}$, we use $Y_{M,M}$ to denote the evaluations of the surrogate model at the M training samples.

5.1.1 Conservatively Estimating a Single Risk Measure

In this section, we compare the performance of various methods for constructing surrogate models used to estimate a single risk measure. Specifically, we assess their ability to conservatively estimate AVaR with $p = 0.8$. In the left panel of Figure 6, we depict the linear ($Q = 1$) PCE approximations (22) built from 30 training samples of (25) (with $D = 1$, $a_1 = 1$) using least squares (LstSq), FSD, SSD and quantile regression (Qnt) with $p = 0.8$. The methods described in this paper introduce a bias to ensure that they conservatively estimate risk. However, it is not just the coefficient θ_0 of the constant term in the polynomial expansion that is affected by the constraints. Higher-order coefficients are influenced as well. This is evident by the different slopes of the linear approximations. In contrast, the safety margin (LstSqB) approach for constructing conservative surrogate models in [35] introduces a constant bias to the least-squares estimator. Specifically, given a pre-defined level of conservativeness $0 < c < 1$, they use leave-one-out cross validation to choose the safety margin s (which is added to θ_0) via $s = -F_{CV}^{-1}(1-c)$, where F_{CV} is the CDF of the cross-validation error. It is unclear from [35] and similar references how to choose c. Here, we set $c = p$.

The two rightmost panels of Figure 6 can be used to assess the level of conservativeness of each surrogate model. For each method, we construct a set of surrogate models using 100 realizations of training data, each consisting of 30 samples from (25) with $D = 3$ and $a_1 = a_2 = a_3 = 1$. The box plots depict the variation in the signed difference between each surrogate and reference values for the risk measure. The lower and upper bounds of the boxes represent the first and third quartiles of the differences, the horizontal line inside the box represents the median, and the lower and upper whiskers represent the minimum and maximum values. The percentage below the lower whisker of each box plot denotes the portion of surrogate models that did not conservatively estimate AVaR using Y_M (middle) and Y_∞ (right). Because the least-squares surrogate model, even with the safety margin modification, is not tuned to estimate AVaR, the estimates of the risk measure are not always conservative with respect to the values obtained from the empirical distribution Y_M or the reference distribution Y_∞. The least-squares surrogate models underestimate the empirical and exact risk measures 96% and 86% of the time, respectively. The bias introduced by the safety margin approach improves these rates. However, the degree of conservatism depends on the hyper-parameter c, which is not easily connected to the target risk measure. In contrast, the FSD, SSD and quantile surrogate models always conservatively estimate $\mathcal{R}(Y_M)$. As expected, they are not always conservative with respect to $\mathcal{R}(Y_\infty)$.

It is impossible to use risk-adapted regression to ensure conservative estimates of risk measures, when training is performed with a finite amount of data. However, increasing conservatism decreases the rate of underestimation of the exact risk. That is, the FSD surrogate model has the lowest rate of underestimation, followed by SSD, and quantile regression. The increased conservatism tends to increase the magnitude of the error in the risk measure. Consequently, if a single risk measure can be elicited from the stakeholder, then the associated regression problem derived from the risk quadrangle should be used to construct the surrogate model. Moreover, the regression problems based on the risk quadrangle are often computationally easier to solve than their stochastic dominance counterparts.

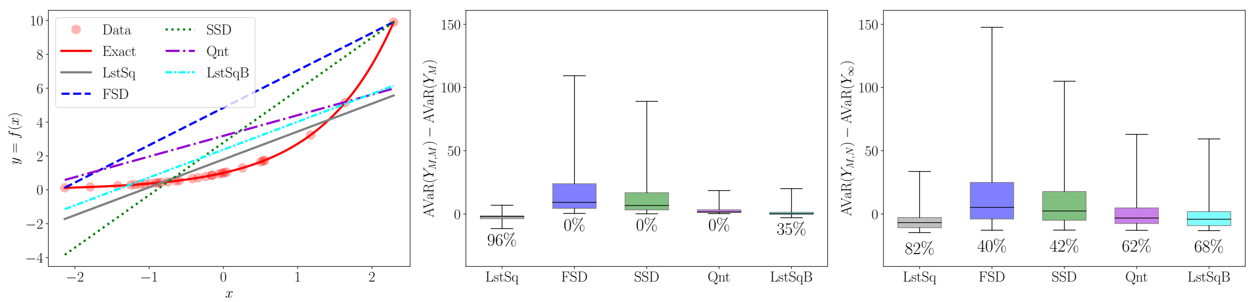

Figure 6: (Left) Linear polynomial approximations for (25) computed using various estimation techniques. (Middle) The signed difference between the surrogate and empirical AVaR estimates computed using the training data for $p = 0.8$ and $M = 30$. (Right) The signed difference between the exact AVaR and the value obtained by evaluating the surrogate at the validation samples. Here, $\mathrm{AVaR}(Y_\infty) \approx 18.23$ and the percentage below the lower whisker of each box plot represents the portion of surrogate models that did not conservatively estimate AVaR.

5.1.2 Conservatively Estimating a Set of Risk Measures

In this section, we demonstrate the efficacy of using stochastic dominance to construct surrogate models that conservatively estimate a set of risk measures. In Figure 7, we construct FSD and SSD quadratic ($Q = 2$) PCE surrogate models using 30 evaluations of (25) with $D = 1$, $a_1 = 1$. The left panel shows that the CDF, computed using the FSD surrogate model, evaluated at the training data (labeled "FSD Train"), always lies to the right of the empirical estimate obtained using the same data (labeled "Train"). This means that the surrogate CDF overestimates the PoF for any threshold as desired.

The left panel of Figure 7 also plots the exact CDF of Y_∞ and the CDF computed using $N = 10^5$ QMC samples Y_N of the FSD surrogate model (labeled "FSD Validation"). The FSD CDF evaluated at the validation samples is still conservative with respect to the empirical estimate, but it is not conservative with respect to the exact CDF over the entire support of Y. Stochastic dominance, indeed any constraint, can only be enforced on the training data. When using 30 training samples, the FSD CDF does not conservatively predict the tail for CDF values greater than approximately 0.96, which can be improved by adding more samples.

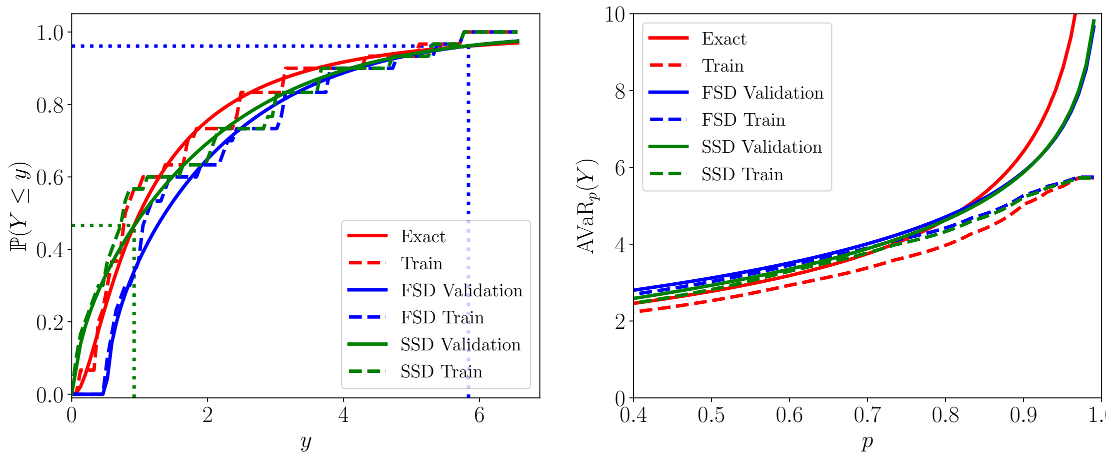

Figure 7: Comparison of the CDFs (left) and AVaRs (right) computed for a range of thresholds s obtained analytically (Exact), empirically using $M = 30$ training samples (Train), using the FSD/SSD approximations evaluated at the training data (FSD/SSD Train), and using the FSD/SSD approximation evaluated at the validation data (FSD/SSD Validation). The CDF estimated by evaluating the FSD surrogate model at the validation sample is not conservative to the right of the vertical dotted blue line. This occurs at $y = 5.73$ such that $\mathbb{P}(Y_N \le y) = 0.96$. The SSD CDF $\mathbb{P}(Y_N \le y)$ is also not conservative at a similar value of y, but is also not conservative to the left of the vertical dotted green line.

As discussed in Section 2.4, the SSD constraints cannot guarantee that the resulting surrogate model conservatively estimates the PoF. However, the resulting surrogate model is still more conservative than the least-squares surrogate models and the risk-quadrangle-based surrogate models described in Section 3.

The right of Figure 7 shows that the values of $\mathrm{AVaR}_p(Y_{M,M})$, computed using the SSD surrogate model evaluated at the training data, is larger than the empirical estimate obtained using the same data for all p. The estimates of $\mathrm{AVaR}_p(Y_{M,N})$ are also closer to the to the reference value $\mathrm{AVaR}_p(Y_\infty)$ when the surrogate model is evaluated at the validation samples.[4] However, the 30 samples do not contain enough extreme values to ensure that the surrogate model is conservative with respect to the exact function. Similar to the FSD surrogate model, the conservativeness of the SSD surrogate model can be improved by using more data and increasing the polynomial degree. As expected, the FSD estimates of AVaR using the empirical and validation data are more conservative than the SSD estimates.

Figure 8 compares the accuracy of different surrogate-based estimates of the PoF with $D = 3$ and $a_1 = a_2 = a_3 = 1$. We define failure to occur when $Y_\infty > C$. We set the threshold to $C = q_{0.95}(Y_\infty)$ so that $\mathbb{P}(Y_\infty > C) = 0.95$. The box plots in the left panel of Figure 8 depict the variation (from 100 realizations of the training data[5]) in the signed difference between the surrogate PoF, $\mathbb{P}(Y_{M,M} > C)$, and the empirical PoF, $\mathbb{P}(Y_M > C)$. For this example, the least-squares surrogate model underestimates the empirical PoF computed using the training data for 11% of the trials, whereas the FSD surrogate model always overestimates the PoF. The SSD and quantile surrogate models perform similarly to FSD for this example, but this is not always the case. The middle panel of Figure 8 plots the accuracy of the PoF estimated using $N = 10^5$ evaluations of each surrogate type. No method is consistently conservative, although least squares is the least conservative. For a fixed PCE degree, the conservatism of all models can be improved by using more evaluations. The improved accuracy in the estimated PoF using surrogate models built with $M = 100$ samples are shown in the right panel of Figure 8.

[4]We only plot $p \in [0.4, 1)$ to better visualize the differences between the different surrogate models. Both FSD and SSD surrogate models conservatively estimate AVaR for $p \in [0, 0.4)$.

[5]97 of the 100 FSD estimation problems converged using a fixed set of optimization hyper-parameters. The failing problems can be solved by hand tuning the solver hyper-parameters. We list the failed run here to demonstrate that solving the FSD-constrained least-square problem is typically more numerically challenging than solving regression problems admitted by the risk quadrangle. Future work is needed to develop optimization algorithms that can more robustly solve these problems without the need to fine tune solver hyper-parameters. We note that algorithm robustness increases as ϵ increases.

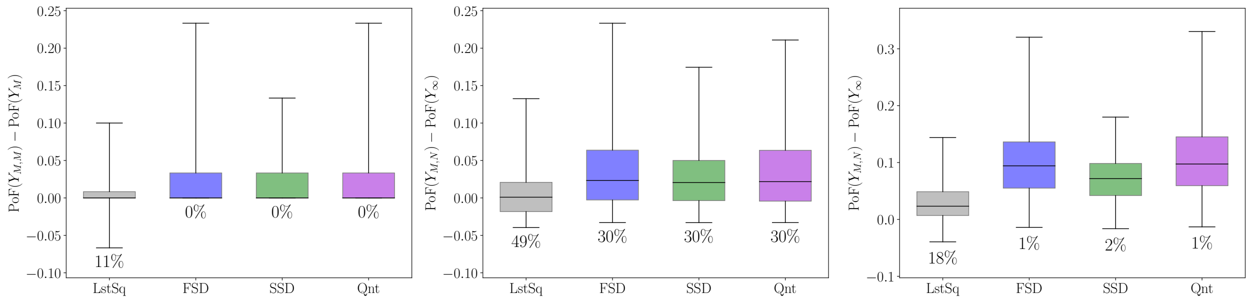

Figure 8: Comparison of the accuracy of different surrogate models when used to compute probability of failure (PoF) $\mathbb{P}(Y > q_{0.95}(Y))$. (Left) The signed difference between the surrogate and empirical PoF computed using 30 training samples. The signed difference between the exact PoF and the PoF obtained by evaluating the surrogate at the validation samples when (middle) $M = 30$ (right) $M = 100$. Here $\text{PoF}(Y_N) = 0.95$. The percentage below the lower whisker of each box plot represents the portion of surrogate models that did not conservatively estimate AVaR.

5.1.3 Convergence

In the previous section, we demonstrated the benefit of introducing a nonconstant bias to ensure that the resulting surrogate model conservatively estimates a fixed risk measure. In this section, we provide numerical evidence to support the theory presented in Sections 3.2 and 4.1, which states that the error in the estimates of the target risk measure decays at the same rate as the mean-squared error. Specifically, we show that as the degree of the PCE increases the estimates of the risk measure, associated with the analytical test function (25), decay exponentially fast.

In Figure 9, we set $a_i = \exp(-2(i-1))$ for $i = 1, 2$ and plot the error in estimates of $\text{AVaR}_{0.9}(Y_\infty) \approx 6.53$ and $\mathbb{P}(Y_\infty > 3.6445) = 0.1$ obtained using different types of risk-adapted surrogate models as the polynomial degree increases. The number of samples was set to increase with degree. Specifically, the number of samples was set to 5 times the number of terms in the PCE. For both quantities, each surrogate model produces an estimate that converges exponentially fast. However, as shown in the previous examples the least-squares surrogate models often do not conservatively estimate the risk measure.

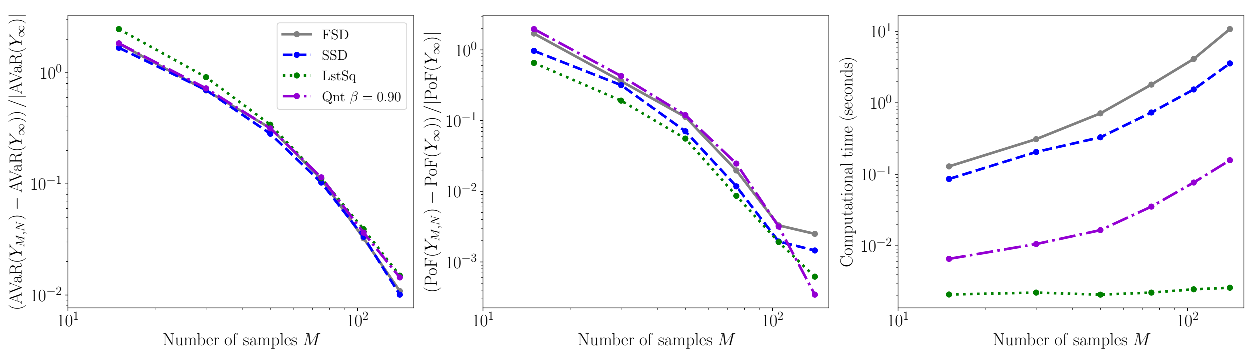

Figure 9: Median error (over 100 trials) in the estimate as a function of the polynomial degree for (left) $\text{AVaR}_{0.9}(Y)$ and (middle) $\mathbb{P}(Y \geq 3.6445)$ obtained using accurate quadrature applied to the surrogate model. Here, $\text{PoF}(Y_\infty)/\text{AVaR}(Y_\infty)$ denotes the value computed analytically and $\text{PoF}(Y_{M,N})/\text{AVaR}(Y_{M,N})$ is estimate using $N = 10^5$ random evaluations of the surrogate model. The median computational time to solve the different regression problems is plotted in the right panel.

The exponential convergence rate depicted in Figure 9 is expected. Propositions 3.1 and 4.1 state that the error in the risk measure obtained by surrogate models based on the risk quadrangle and stochastic dominance, respectively, decays at a similar rate as the truncation error of the PCE. For analytical functions, such as (25) used here, the truncation error decays exponentially fast. For functions of the solution to a partial differential equation, we can employ the results in, e.g., [59, 60] to determine the anticipated convergence rate.

The computational time needed to solve the quantile, FSD and SSD regression problems depends on the amount of training data and the number of unknowns in the surrogate model. The right panel of Figure 9 plots the median computational time required to solve each regression problem. The optimization problems used to construct conservative surrogate models are more expensive than least squares regression. However, our paper presents methods for estimating risk measures from limited data, assuming that the cost of evaluating the numerical model used to generate the data is computationally expensive. In this setting the cost of solving the optimization problem is negligible to the cost of collecting data.

5.2 Risk Estimation for a Truss Structure

In this section, we investigate the utility of risk-adapted surrogate models for a more realistic engineering benchmark. Specifically, we consider the two-dimensional truss structure depicted in Figure 1. The model of this structure is parameterized by 10 independent random variables representing the uncertain loads and material properties. The Young's moduli, E_1 and E_2 (Pa), are log-normally distributed with mean 2.1×10^{11} and standard deviation 2.1×10^{10}. The cross-sectional area of the horizontal trusses A_1 (m^2) is log-normally distributed with mean 2.0×10^{-3} and standard deviation 2.0×10^{-4}, and the cross-sectional area of the diagonal trusses A_2 (m^2) is log-normally distributed with mean 10^{-3} and standard deviation 10^{-4}. The loads $P_1,\ldots,P_6$ (N) satisfy a Gumbel distribution with mean 5.0×10^4 and standard deviation 7.5×10^3. This problem was used in [61] and has been used to benchmark methods for estimating rare events in [17, 62]. The authors of these works provided a data-set containing the vertical displacement y at the mid-span for 10,000 random MC samples of the uncertain parameters.[6] We used this data for training and validation of risk-adapted PCE surrogate models. We use the Nataf transformation [43] to transform the log-normal variables into standard Normal variables and use Hermite polynomials as the univariate basis for each of these transformed variables.[7] We numerically generate the basis associated with the Gumbel variables using the predictor-corrector algorithm in [55].

Figure 10 compares the reference $\mathrm{AVaR}_{0.9}(Y_N)\approx0.1010$ with the estimates obtained from quadratic PCE models ($K=66$) constructed with four different regression methods each using 100 different realizations of 200 training samples. We use 9,000 validation samples, independent of the training samples, to estimate the reference AVaR. The FSD, SSD and quantile surrogate models always conservatively estimate AVaR with respect to the training data, but only FSD and SSD overestimate AVaR with respect to the validation data. The least-squares model almost always underestimates AVaR. The FSD and SSD model also perform well when estimating the PoF $\mathbb{P}(Y_N>q_p(Y_N))=0.1$, where $q_p(Y_N)\approx0.9419$, computed using $M=100$ training data and the validation data. See the left and right plots of Figure 11, respectively.[8] However, quantile regression performs almost as poorly as basic least-squares regression when estimating the validation PoF because it no longer targets the correct statistic.

5.3 Wing-weight Model

In this section, we show that our framework for constructing surrogate models can be applied to approximation strategies other than traditional polynomial regression used in the previous examples. Specifically, we show that it can be coupled with dimension reduction to significantly reduce the cost of risk estimation in high-dimensions. Dimension reduction has been used to improve the accuracy of surrogate estimates of risk in [64, 65], but these methods do not guarantee conservative estimates of the risk measure.

We consider the model for wing-weight used in [66], i.e.,

$$f(x)=0.036\,S_w^{0.758}W_{fw}^{0.0035}\left(\frac{A}{\cos^2(\Lambda)}\right)^{0.6}q^{0.006}\lambda^{0.04}\left(\frac{100t_c}{\cos(\Lambda)}\right)^{-0.3}(N_zW_{dg})^{0.49}+S_wW_p, \tag{26}$$

where $x=[S_w,W_{fw},A,\Lambda,q,\lambda,t_c,N_z,W_{dg},W_p]^\top$. The distributions of the random parameters are provided in Table 2. In the following, we reduce the dimensionality by finding important directions of the input

[6]The data can be obtained from `https://www.uqlab.com`.

[7]The Nataf transformation is necessary because the polynomials orthonormal to the log-normal distribution do not form a basis [63].

[8]FSD fails to conservatively estimate the PoF for one realization of the training data. This is a numerical artifact. The PoF is only underestimated by approximately 10^{-5}. This error is due to solving the FSD problem to a fixed, non-zero accuracy, constrained through optimization tolerances and because of the use of the smoothing of the max function. We report this result to make sure that the readers are aware of this possible phenomenon.

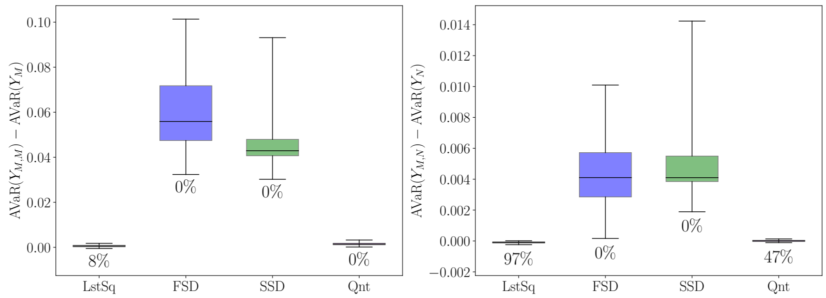

Figure 10: The signed difference between the surrogate and empirical estimate of $AVaR_p$, $p = 0.9$, using the (left) $M = 200$ training data and (right) the validation data. validation data. Here, $\mathrm{AVaR}_{0.9}(Y_N) \approx 0.1010$ is computed analytically, The percentage below the lower whisker of each box plot represents the portion of surrogate models that did not conservatively estimate AVaR.

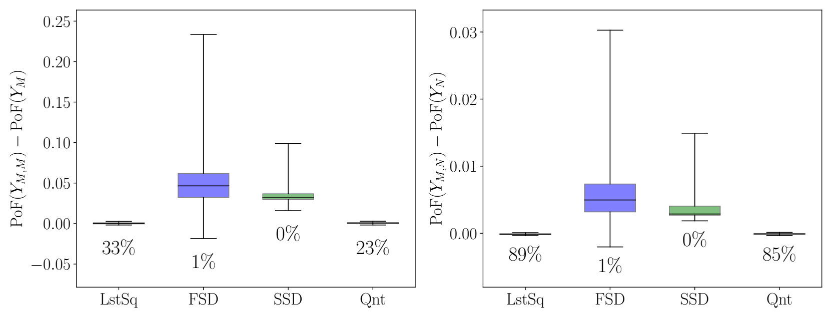

Figure 11: The signed difference between the surrogate and empirical PoF computed using (left) the $M = 100$ training data and (right) the validation data. Here, $\mathbb{P}(Y_N > 0.9419) = 0.1$ is computed analytically, $\mathrm{PoF}(Y_M)$ is computed using the M training samples, $\mathrm{PoF}(Y_{M,M})$ is estimated using M evaluations of the surrogate at the training samples, and $\mathrm{PoF}(Y_{M,N})$ is estimated using $N = 10^5$ QMC evaluations of the surrogate model. The percentage below the lower whisker of each box plot represents the portion of surrogate models that did not conservatively estimate PoF.

space using active subspaces [67]. We determine the important directions by computing the eigenvalue decomposition of the average outer-products of the gradients of the model f at a set of M samples, that is

$$WDW^\top = C = \frac{1}{M}\sum_{m=1}^{M} \nabla_x f(x^{(m)}) \nabla_x f(x^{(m)})^\top$$

Figure 12 plots the projection of $M = 30$ MC evaluations of f onto the dominant one-dimensional (1D) subspace, defined by the first eigenvector of W. It is evident that the wing-weight model can be reasonably approximated by a 1D function. However, no 1D approximation can be exact due to small variations in the model response in the inactive directions. Even so, we can still accurately and conservatively approximate risk using the framework presented in this paper. In Figure 12, we compare the quadratic FSD, SSD and quantile ($p = 0.9$) surrogate models built on the dominant 1D active subspace using the 30 training samples depicted. Each surrogate model conservatively and accurately approximates the empirical estimate of AVaR for $p = 0.9$. Specifically, the signed relative differences $(\mathrm{AVaR}_{0.9}(Y_{M,M}) - \mathrm{AVaR}_{0.9}(Y_M))/\mathrm{AVaR}_{0.9}(Y_m)$ are 0.023 (FSD), 0.002 (SSD), and 0.005 (Quantile). In contrast the least-squares surrogate model does not conservatively estimate risk; the associated signed relative error is -0.002.

6 Conclusions

We presented two techniques for conservatively estimating (overestimating) risk measures using a fixed data set. Such situations arise when archived data is available and there is no recourse to further interrogate the

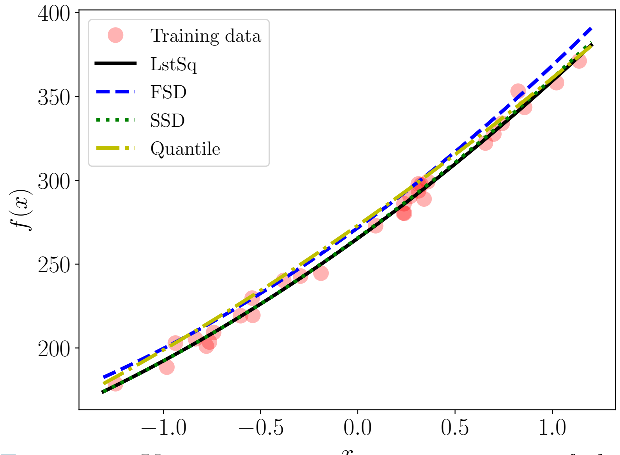

Figure 12: Various surrogate approximations of the wing-weight model (26).

Table 2: Random variables of the wing weight model (26)

Variable	Symbol	Distribution
Wing area	S_w	$\mathcal{U}(150, 200)$
Weight of fuel in the wing	W_{fw}	$\mathcal{U}(220, 300)$
Aspect ratio	A	$\mathcal{U}(6, 10)$
Quarter-chord sweep	Λ	$\mathcal{U}(-5\pi/9, 5\pi/9)$
Dynamic pressure at cruise	q	$\mathcal{U}(16, 45)$
Taper ratio	λ	$\mathcal{U}(0.5, 1)$
Aerofoil thickness to chord ratio	t_c	$\mathcal{U}(0.08, 0.18)$
Ultimate load factor	N_z	$\mathcal{U}(2.5, 6)$
Flight design gross weight	W_{dg}	$\mathcal{U}(1700, 2500)$
Paint weight	W_p	$\mathcal{U}(0.025, 0.08)$

data source. The first approach uses the risk quadrangle to construct surrogate models that conservatively estimate a risk measure associated with the subjective utility/regret preferences of a stakeholder. The second approach uses stochastic dominance to ensure that our estimates are conservative with respect to a set of risk measures, when a single utility preference cannot be agreed upon. For both approaches, we derived theoretical upper bounds on the error for orthonormal surrogate models. The error was shown to converge at a similar rate as the root-mean-squared error when using, e.g., a polynomial chaos expansions. We confirmed these results with numerical examples that showed that risk-adapted surrogate models do indeed overestimate risk, while converging at the expected rate.

Constructing surrogate models with a large number of inputs is challenging. Consequently, surrogate modeling techniques are often built on lower dimensional subspaces identified by dimension-reduction techniques. However, the use of dimension reduction introduces an error that is difficult to quantify. Through a numerical example, we demonstrated that the methods proposed in this paper can be used to conservatively estimate risk even in the presence of such error.

Although not considered in this paper, the approximation strategy presented here can also be used within an adaptive sampling framework to generate samples that can improve the surrogate model's estimate of risk. Existing adaptive strategies build a surrogate model from available training data and use properties of the surrogate model to select the new data. Although these sampling strategies can target risk measures [28, 29, 68], they are typically restricted to reduced order modeling for partial differential equations, not regression, and therefore do not directly apply to our methods. We will explore the extension of such adaptive sampling strategies to our conservative estimation procedures in future work.

Acknowledgments

This work was sponsored by the Sandia National Laboratories Laboratory Directed Research Development (LDRD) program and the US Department of Energy, Office of Advanced Scientific Computing Research.
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.

The authors would like to thank Joseph Hart and Bart Van Bloemen Waanders at Sandia National Laboratories for this insightful comments made on initial drafts of this manuscript.

Appendix A Proof of conservatism (7) and Proposition 3.1

In this section, we demonstrate that the two step procedure in equations (5) and (6) produces a surrogate model that conservatively estimates a predefined risk measure. We then prove Proposition 3.1. As seen in Figure 3 an error measure defines a deviation measure, which allows us to rewrite the regression problem (5) in the equivalent, two-step, form

$$\theta \in \arg\min_{\theta' \in \Theta} \mathcal{D}(Y - g(X, \theta')) \quad \text{with} \quad 0 \in \mathcal{S}(Y - \theta_0 - g(X, \theta)). \tag{27}$$

This form provides insight that is not immediately evident from (5). Specifically, the optimal model parameters θ minimize the deviation or uncertainty associated with the model. Moreover, the statistic of the residual must be zero or equivalently, the optimal θ_0 is a statistic of the residual $\theta_0 \in \mathcal{S}(Y - g(X, \theta))$. For least-squares regression, the optimal θ minimize the standard deviation of the residual and the associated statistic, the average of the residual, $\mathbb{E}[Y - \theta_0 - g(X, \theta)]$, is zero. Consequently, least-squares regression is inappropriate when large positive outcomes are undesirable. In contrast, the statistic of the quantile risk quadrangle is the p-th quantile of the residual, which is more representative of undesirable events. The decomposition of θ and θ_0 in (27) may also provide computational benefits depending on the numerical properties of the deviation measure $\mathcal{D}$.

To construct risk-adapted surrogate models that conservatively estimate risk, in the sense that they try to avoid underestimation, we use the following result. We note that this result extends Corollary 4.2 in [36] with the addition of a lower bound based on the expectation of the surrogate model.

Proposition Appendix A.1. *Suppose $\mathcal{R}$ is positively homogeneous and that there exists $\bar{\theta} \in \Theta$ for which $g(x, \bar{\theta}) = 0$ for all $x \in \Omega_X$. If $\theta^\star \in \Theta$ solves the deviation minimization problem*

$$\min_{\theta \in \Theta} \mathcal{D}(Y - g(X, \theta)), \tag{28}$$

and we set

$$\theta_0^\star := \mathcal{R}(Y - g(X, \theta^\star)), \tag{29}$$

then the following bounds hold

$$\theta_0^\star + \mathbb{E}[g(X, \theta^\star)] \leq \mathcal{R}(Y) \leq \theta_0^\star + \mathcal{R}(g(X, \theta^\star)) = \theta_0^\star + \mathbb{E}[g(X, \theta^\star)] + \mathcal{D}(g(X, \theta^\star)). \tag{30}$$

Proof[Proof of Proposition Appendix A.1] As a consequence of the relationships depicted in Figure 3, any risk measure $\mathcal{R}$ generated by the risk quadrangle is translation equivariant (R3). The translation equivariance of $\mathcal{R}$ ensures that $\mathcal{R}(\theta_0 + g(X, \theta)) = \theta_0 + \mathcal{R}(g(X, \theta))$ since the shift parameter θ_0 is constant. If $\mathcal{R}$ is positively homogeneous (R4), then so are $\mathcal{D}$, $\mathcal{V}$ and $\mathcal{E}$. Moreover, positive homogeneity combined with convexity ensures that $\mathcal{R}$ is in fact subadditive and consequently,

$$\begin{aligned} \mathcal{R}(Y) &= \mathcal{R}(\theta_0 + g(X, \theta) + (Y - \theta_0 - g(X, \theta))) \\ &\leq \mathcal{R}(\theta_0 + g(X, \theta)) + \mathcal{R}(Y - \theta_0 - g(X, \theta)). \end{aligned} \tag{31}$$

By choosing $\theta_0 = \mathcal{R}(Y - g(X, \theta))$, a straightforward consequence of (31) and the translation equivariance of $\mathcal{R}$ is that $\mathcal{R}(\theta_0 + g(X, \theta)) \geq \mathcal{R}(Y)$. Moreover, this choice of θ_0 ensures that the expectation of the model satisfies

$$\mathbb{E}[\theta_0 + g(X, \theta)] = \mathbb{E}[Y] + \mathcal{D}(Y - g(X, \theta)). \tag{32}$$

The upper bound in (30) follows from (31) and our choice of $\theta_0^\star$. Specifically, we have that

$$\mathcal{R}(Y) \leq \mathcal{R}(\theta_0^\star + g(X, \theta^\star)) + \mathcal{R}(Y - \theta_0^\star - g(X, \theta^\star) = \theta_0^\star + \mathcal{R}(g(X, \theta^\star)),$$

where we have used the facts that $\mathcal{R}(\theta_0^\star + g(X, \theta^\star)) = \theta_0^\star + \mathcal{R}(g(X, \theta^\star))$ and $\mathcal{R}(Y - \theta_0^\star - g(X, \theta^\star)) = 0$. On the other hand, the assumption that there exists $\bar{\theta}$ for which $g(x, \bar{\theta}) = 0$ for all $x \in \Omega_X$ and the optimality of $\theta^\star$ ensure that

$$\mathcal{R}(Y) = \mathbb{E}[Y] + \mathcal{D}(Y) = \mathbb{E}[Y] + \mathcal{D}(Y - g(X, \bar{\theta})) \geq \mathbb{E}[Y] + \mathcal{D}(Y - g(X, \theta^\star)).$$

Therefore, (32) and the translation equivariance of $\mathcal{R}$ yield the lower bound in (30). □

We note that if Y is exactly represented by the model $\theta_0 + g(X,\theta)$, the error minimization problem will return $\theta_0' \in \mathbb{R}$ and $\theta^\star \in \Theta$ that produce zero error, i.e., $Y = \theta_0' + g(X,\theta^\star)$. Likewise, $\theta^\star$ will produce zero deviation $\mathcal{D}(Y - g(X,\theta^\star)) = 0$. Therefore,

$$\theta_0^\star + \mathbb{E}[g(X,\theta^\star)] = \mathbb{E}[Y] = \theta_0' + \mathbb{E}[g(X,\theta^\star)] \quad \Longrightarrow \quad \theta_0^\star = \theta_0'$$

and hence the above estimation procedures produce $\theta_0^\star$ and $\theta^\star$ that exactly reproduce Y.

Remark Appendix A.1 (Deviation Error Bounds)**.** *An important consequence of* (30) *in Proposition Appendix A.1 is that the average of the surrogate model $\theta_0^\star + \mathbb{E}[g(X,\theta^\star)]$ underestimates the risk by at most the deviation $\mathcal{D}(g(X,\theta^\star))$, i.e.,*

$$0 \le \mathcal{R}(Y) - (\theta_0^\star + \mathbb{E}[g(X,\theta^\star)]) \le \mathcal{D}(g(X,\theta^\star)), \tag{33}$$

which follows from the risk quadrangle relationships depicted in Figure 3. Similarly, the risk of the surrogate model overestimates the true risk again by $\mathcal{D}(g(X,\theta^\star))$, i.e.,

$$0 \le (\theta_0^\star + \mathcal{R}(g(X,\theta^\star))) - \mathcal{R}(Y) \le \mathcal{D}(g(X,\theta^\star)). \tag{34}$$

Remark Appendix A.2 (Choice of Shift Parameter)**.** *Although the deviation minimization problem* (28) *is intuitive—we seek parameters $\theta^\star$ that minimize the uncertainty associated with the residual—the numerical solution of* (28) *can be more challenging than the error minimization problem* (5)*. Fortunately, the risk quadrangle relationships depicted in Figure 3 ensure that the optimal $\theta^\star$ in Proposition Appendix A.1 also solves the error minimization problem* (5)*. As a consequence, we can solve* (5) *and still ensure that the resulting surrogate model conservatively estimates $\mathcal{R}(Y)$ by computing the shift* (29) *from Proposition Appendix A.1.*[9]

Remark Appendix A.3 (Positive Homogeneity)**.** *If $\mathcal{R}$ is not positively homogeneous, we can modify* (31) *as*

$$\mathcal{R}(Y) \le \mathcal{R}_\lambda(\theta_0 + g(X,\theta)) + \mathcal{R}_{1-\lambda}(Y - \theta_0 - g(X,\theta))$$

for any $\lambda \in (0,1)$, where $\mathcal{R}_t(X) := t\mathcal{R}(X/t)$ for $t > 0$. Note that $\mathcal{R}_t$ is associated with a risk quadrangle with components $\mathcal{E}_t$, $\mathcal{D}_t$ and $\mathcal{V}_t$ defined analogously. In this setting, the minimum deviation upper bound in (30) *holds with $\mathcal{R}$ replaced by $\mathcal{R}_{1-\lambda}$. However, the lower bound in* (30) *may no longer hold. To circumvent this, one may consider replacing the error measure used in* (5) *with either*

$$\mathcal{E}^1(X) := \inf_{0\le t\le 1} \mathcal{E}_t(X) \quad \text{or} \quad \mathcal{E}^2(X) := \inf_{t\ge 0} \mathcal{E}_t(X).$$

Unfortunately, the first error measure is equal to $\mathcal{E}$ and the second is identically zero for many practical $\mathcal{E}$. As a simple example, consider the mean-squared error $\mathcal{E}(X) = \frac{1}{2}\mathbb{E}[X^2]$. It follows that

$$\mathcal{E}_t(X) = \tfrac{1}{2t}\mathbb{E}[X^2],$$

which is strictly decreasing with respect to t and hence $\mathcal{E}^1 = \mathcal{E}$ and $\mathcal{E}^2 \equiv 0$.

This brings us to the proof of Proposition 3.1.

Proof[Proof of Proposition 3.1] We denote the first K Fourier coefficients by $\bar\theta_{1:K}$. The optimality of $\theta^\star$ and the fundamental risk quadrangle relationships ensure that

$$\begin{aligned}
\mathcal{D}(Y - g_K(X,\theta^\star)) + &\mathcal{D}(g_K(X,\theta^\star)) - \mathcal{D}(Y) \\
&\le \mathcal{D}(Y - g_K(X,\bar\theta_{1:K})) + \mathcal{D}(g_K(X,\bar\theta_{1:K})) - \mathcal{D}(Y) \\
&\le \mathcal{D}(Y - g_K(X,\bar\theta_{1:K})) + \mathcal{D}(g_K(X,\bar\theta_{1:K}) - Y) \\
&\le \mathcal{E}(Y - g_K(X,\bar\theta_{1:K})) + \mathcal{E}(g_K(X,\bar\theta_{1:K}) - Y) \\
&\le 2C\tau_K^\alpha
\end{aligned}$$

as was to be proved. Here, the first inequality follows from the optimality of $\theta^\star$, the second follows from (10), the third follows from the definition of $\mathcal{D}$, and the fourth follows from (12). □

[9]The optimal shift θ_0' from (5) satisfies $\theta_0' \in \mathcal{S}(Y - g(X,\theta^\star))$, but need not coincide with $\theta_0^\star$ unless $\mathcal{S}(\cdot) = \mathcal{R}(\cdot)$. This is the case, for example, with the entropic risk quadrangle for which $\mathcal{R}$ is the entropic risk measure, $\mathcal{R}(Y) = \ln \mathbb{E}[\exp(Y)]$. However, the entropic risk measure is not positively homogeneous.

Appendix B Proof of Proposition 4.1

The proof of Proposition 4.1 proceeds as follows.

Proof[Proof of Proposition 4.1] Note that we have the following inequalities:

$$\begin{aligned}\tau_K = \mathbb{E}[(Y - \bar{\theta}_0 - g_K(X, \bar{\theta}_{1:K}))^2] &\le \mathbb{E}[(Y - \theta_0^\star - g_K(X, \theta^\star))^2] \\ &\le \mathbb{E}[(Y - \theta_0 - g_K(X, \theta))^2],\end{aligned} \tag{35}$$

for all $(\theta_0, \theta) \in \mathbb{R} \times \Theta_K$ satisfying the stochastic dominance constraints for $t \in S$. We now find a $\theta_0 \in \mathbb{R}$ such that $(\theta_0, \bar{\theta}_{1:K})$ is a feasible point. To this end, we denote by r_K the remainder associated with the K-th approximation, i.e., $r_K(x) = f(x) - \bar{\theta}_0 - g_K(x, \bar{\theta}_{1:K})$, and choose

$$\theta_0 = \bar{\theta}_0 + \sup_{x \in f^{-1}(S)} r_K(x).$$

Consequently, we have that

$$f(x) = \bar{\theta}_0 + g_K(X, \bar{\theta}_{1:K}) + r_K(x) \le \theta_0 + g_K(X, \bar{\theta}_{1:K}) \quad \forall\, x \in f^{-1}(S).$$

Therefore, $(\theta_0, \bar{\theta}_{1:K})$ satisfies the FSD and SSD constraints and thus (35) ensures that

$$\begin{aligned}\tau_k \le \mathbb{E}[(Y - \theta_0^\star - g_K(X, \theta^\star))^2] &\le \mathbb{E}[(Y - \theta_0 - g_K(X, \bar{\theta}_{1:K}))^2] \\ &= \tau_k + \left(\sup_{x \in f^{-1}(S)} r_K(x)\right)^2 \le \tau_K + \delta_K^2.\end{aligned}$$

We note here that since f is continuous and S is compact, then $f^{-1}(S)$ is compact and the supremum of r_k over $f^{-1}(S)$ exists and is finite. Expanding the quadratic objective function and utilizing the orthonormality of ψ_ℓ yields the desired result. □

References

[1] E. Zio, The future of risk assessment, Reliability Engineering & System Safety 177 (2018) 176–190. doi:https://doi.org/10.1016/j.ress.2018.04.020.

[2] A. Shapiro, D. Dentcheva, A. Ruszczynski, Lectures on Stochastic Programming: Modeling and Theory, Second Edition, MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics, Philadelphia, 2014.

[3] L. Bonfiglio, J. O. Royset, Multidisciplinary risk-adaptive set-based design of supercavitating hydrofoils, AIAA Journal 57 (8) (2019) 3360–3378. https://doi.org/10.2514/1.J057731, doi:10.2514/1.J057731.
URL https://doi.org/10.2514/1.J057731

[4] A. Chaudhuri, B. Kramer, M. Norton, J. O. Royset, K. Willcox, Certifiable risk-based engineering design optimization, AIAA Journal. In press (2021). arXiv:2101.05129.

[5] D. P. Kouri, A. Shapiro, Optimization of PDEs with Uncertain Inputs, Springer New York, New York, NY, 2018, pp. 41–81. doi:10.1007/978-1-4939-8636-1_2.

[6] D. P. Kouri, T. M. Surowiec, Existence and optimality conditions for risk-averse PDE-constrained optimization, SIAM/ASA Journal on Uncertainty Quantification 6 (2) (2018) 787–815.

[7] D. P. Kouri, T. M. Surowiec, Risk-averse PDE-constrained optimization using the conditional value-at-risk, SIAM Journal on Optimization 26 (1) (2016) 365–396. doi:10.1137/140954556.

[8] J. O. Royset, L. Bonfiglio, G. Vernengo, S. Brizzolara, Risk-Adaptive Set-Based Design and Applications to Shaping a Hydrofoil, Journal of Mechanical Design 139 (10), 101403 (08 2017). https://asmedigitalcollection.asme.org/mechanicaldesign/article-pdf/139/10/101403/6401436/md_139_10_101403.pdf, doi:10.1115/1.4037623.
URL https://doi.org/10.1115/1.4037623

[9] Q. D. Cao, Y. Choe, Cross-entropy based importance sampling for stochastic simulation models, Reliability Engineering & System Safety 191 (2019) 106526. doi:https://doi.org/10.1016/j.ress.2019.106526.
URL https://www.sciencedirect.com/science/article/pii/S0951832018309219

[10] A. Chaudhuri, B. Kramer, K. E. Willcox, Information reuse for importance sampling in reliability-based design optimization, Reliability Engineering & System Safety 201 (2020) 106853. doi:https://doi.org/10.1016/j.ress.2020.106853.
URL https://www.sciencedirect.com/science/article/pii/S0951832019301620

[11] L. J. Hong, Z. Hu, G. Liu, Monte carlo methods for value-at-risk and conditional value-at-risk: A review, ACM Trans. Model. Comput. Simul. 24 (4) (2014) 22:1–22:37. doi:10.1145/2661631.
URL http://doi.acm.org/10.1145/2661631

[12] R. M. Neal, Annealed importance sampling, Statistics and Computing 11 (2) (2001) 125–139. doi:10.1023/A:1008923215028.
URL https://doi.org/10.1023/A:1008923215028

[13] S.-K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (4) (2001) 263 – 277. doi:https://doi.org/10.1016/S0266-8920(01)00019-4.
URL http://www.sciencedirect.com/science/article/pii/S0266892001000194

[14] M. D. Shields, D. G. Giovanis, V. Sundar, Subset simulation for problems with strongly non-gaussian, highly anisotropic, and degenerate distributions, Computers & Structures 245 (2021) 106431. doi:https://doi.org/10.1016/j.compstruc.2020.106431.
URL https://www.sciencedirect.com/science/article/pii/S0045794920302340

[15] Z. Wang, M. Broccardo, J. Song, Hamiltonian monte carlo methods for subset simulation in reliability analysis, Structural Safety 76 (2019) 51–67. doi:https://doi.org/10.1016/j.strusafe.2018.05.005.
URL https://www.sciencedirect.com/science/article/pii/S0167473017301820

[16] H. Lim, L. Manuel, Distribution-free polynomial chaos expansion surrogate models for efficient structural reliability analysis, Reliability Engineering & System Safety 205 (2021) 107256. doi:https://doi.org/10.1016/j.ress.2020.107256.
URL https://www.sciencedirect.com/science/article/pii/S0951832020307560

[17] S. Marelli, B. Sudret, An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis, Structural Safety 75 (2018) 67 – 74. doi:https://doi.org/10.1016/j.strusafe.2018.06.003.
URL http://www.sciencedirect.com/science/article/pii/S0167473017302977

[18] A. J. Torii, R. H. Lopez, L. F. F. Miguel, Probability of failure sensitivity analysis using polynomial expansion, Probabilistic Engineering Mechanics 48 (2017) 76 – 84. doi:https://doi.org/10.1016/j.probengmech.2017.06.001.
URL http://www.sciencedirect.com/science/article/pii/S0266892017301273

[19] B. Bichon, M. Eldred, L. Swiler, S. Mahadevan, J. McFarland, Efficient global reliability analysis for nonlinear implicit performance functions, AIAA Journal 46 (10) (2008) 2459–2468. doi:10.2514/1.34321.

[20] B. Gaspar, A. Teixeira, C. Guedes Soares, Adaptive surrogate model with active refinement combining kriging and a trust region method, Reliability Engineering and System Safety 165 (2017) 277–291, cited By 83. doi:10.1016/j.ress.2017.03.035.
URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85018492783&doi=10.1016%2fj.ress.2017.03.035&partnerID=40&md5=f11204da9d61fd2cd8d9c94560058469

[21] Z. Sun, J. Wang, R. Li, C. Tong, Lif: A new kriging based learning function and its application to structural reliability analysis, Reliability Engineering and System Safety 157 (2017) 152–165, cited By 120. doi:10.1016/j.ress.2016.09.003.
URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84995611280&doi=10.1016%2fj.ress.2016.09.003&partnerID=40&md5=1b22647ea906dfa5df1aed4a9e14db1b

[22] V. S. Sundar, M. D. Shields, Reliability analysis using adaptive kriging surrogates with multimodel inference, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 5 (2) (2019) 04019004. doi:10.1061/AJRUA6.0001005.

[23] S. Xiao, S. Oladyshkin, W. Nowak, Reliability analysis with stratified importance sampling based on adaptive kriging, Reliability Engineering & System Safety 197 (2020) 106852. doi:https://doi.org/10.1016/j.ress.2020.106852.
URL http://www.sciencedirect.com/science/article/pii/S0951832019301383

[24] Y. Bao, Z. Xiang, H. Li, Adaptive subset searching-based deep neural network method for structural reliability analysis, Reliability Engineering & System Safety 213 (2021) 107778. doi:https://doi.org/10.1016/j.ress.2021.107778.
URL https://www.sciencedirect.com/science/article/pii/S0951832021003033

[25] S. Chatterjee, S. Sarkar, S. Hore, N. Dey, A. S. Ashour, V. E. Balas, Particle swarm optimization trained neural network for structural failure prediction of multistoried rc buildings, Neural Computing and Applications 28 (8) (2017) 2005–2016. doi:10.1007/s00521-016-2190-2.
URL https://doi.org/10.1007/s00521-016-2190-2

[26] V. Papadopoulos, D. G. Giovanis, N. D. Lagaros, M. Papadrakakis, Accelerated subset simulation with neural networks for reliability analysis, Computer Methods in Applied Mechanics and Engineering 223-224 (2012) 70 – 80. doi:https://doi.org/10.1016/j.cma.2012.02.013.
URL http://www.sciencedirect.com/science/article/pii/S0045782512000552

[27] M. Heinkenschloss, B. Kramer, T. Takhtaganov, K. Willcox, Conditional-value-at-risk estimation via reduced-order models, SIAM/ASA Journal on Uncertainty Quantification 6 (4) (2018) 1395–1423. https://doi.org/10.1137/17M1160069, doi:10.1137/17M1160069.
URL https://doi.org/10.1137/17M1160069

[28] Z. Zou, D. P. Kouri, W. Aquino, An adaptive sampling approach for solving PDEs with uncertain inputs and evaluating risk, in: 19th AIAA Non-Deterministic Approaches Conference, 2017, p. 1325.

[29] Z. Zou, D. P. Kouri, W. Aquino, An adaptive local reduced basis method for solving PDEs with uncertain inputs and evaluating risk, Computer Methods in Applied Mechanics and Engineering 345 (2019) 302–322.

[30] M. Li, Z. Wang, Surrogate model uncertainty quantification for reliability-based design optimization, Reliability Engineering & System Safety 192 (2019) 106432, complex Systems RAMS Optimization: Methods and Applications. doi:https://doi.org/10.1016/j.ress.2019.03.039.
URL https://www.sciencedirect.com/science/article/pii/S0951832018305611

[31] R. Rocchetta, L. G. Crespo, A scenario optimization approach to reliability-based and risk-based design: Soft-constrained modulation of failure probability bounds, Reliability Engineering & System Safety 216 (2021) 107900. doi:https://doi.org/10.1016/j.ress.2021.107900.
URL https://www.sciencedirect.com/science/article/pii/S095183202100418X

[32] Z. Wang, A. Shafieezadeh, On confidence intervals for failure probability estimates in kriging-based reliability analysis, Reliability Engineering & System Safety 196 (2020) 106758. doi:https://doi.org/10.1016/j.ress.2019.106758.
URL https://www.sciencedirect.com/science/article/pii/S0951832018310858

[33] Q. Ouyang, W. Lu, J. Lin, W. Deng, W. Cheng, Conservative strategy-based ensemble surrogate model for optimal groundwater remediation design at dnapls-contaminated sites, Journal of Contaminant Hydrology 203 (2017) 1–8. doi:https://doi.org/10.1016/j.jconhyd.2017.05.007.
URL https://www.sciencedirect.com/science/article/pii/S0169772216302984

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

[35] F. A. C. Viana, V. Picheny, R. T. Haftka, Using cross validation to design conservative surrogates, AIAA Journal 48 (10) (2010) 2286–2298. https://doi.org/10.2514/1.J050327, doi:10.2514/1.J050327.
URL https://doi.org/10.2514/1.J050327

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

[37] R. T. Rockafellar, S. Uryasev, The fundamental risk quadrangle in risk management, optimization and statistical estimation, Surveys in Operations Research and Management Science 18 (1–2) (2013) 33 – 53. doi:http://dx.doi.org/10.1016/j.sorms.2013.03.001.
URL http://www.sciencedirect.com/science/article/pii/S1876735413000032

[38] D. Dentcheva, A. Ruszczyński, Optimization with stochastic dominance constraints, SIAM Journal on Optimization 14 (2) (2003) 548–566. doi:10.1137/S1052623402420528.
URL http://dx.doi.org/10.1137/S1052623402420528

[39] H. Föllmer, A. Schied, Convex measures of risk and trading constraints, Finance Stoch. 6 (4) (2002) 429–447. doi:10.1007/s007800200072.
URL http://dx.doi.org/10.1007/s007800200072

[40] P. Artzner, F. Delbaen, J.-M. Eber, D. Heath, Coherent measures of risk, Mathematical Finance 9 (3) (1999) 203–228. doi:10.1111/1467-9965.00068.
URL http://dx.doi.org/10.1111/1467-9965.00068

[41] R. T. Rockafellar, J. O. Royset, On buffered failure probability in design and optimization of structures, Reliability Engineering & System Safety 95 (5) (2010) 499 – 510.

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

[43] J. D. Jakeman, F. Franzelin, A. Narayan, M. Eldred, D. Plfüger, Polynomial chaos expansions for dependent random variables, Computer Methods in Applied Mechanics and Engineering 351 (2019) 643 – 666. doi:https://doi.org/10.1016/j.cma.2019.03.049.
URL http://www.sciencedirect.com/science/article/pii/S0045782519301884

[44] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.

[45] A. Doostan, A. Validi, G. Iaccarino, Non-intrusive low-rank separated approximation of high-dimensional stochastic models, Computer Methods in Applied Mechanics and Engineering 263 (2013) 42–55.

[46] A. A. Gorodetsky, J. D. Jakeman, Gradient-based optimization for regression in the functional tensor-train format, Journal of Computational Physics 374 (2018) 1219 – 1238. doi:https://doi.org/10.1016/j.jcp.2018.08.010.
URL http://www.sciencedirect.com/science/article/pii/S0021999118305321

[47] E. Haber, L. Ruthotto, Stable architectures for deep neural networks, Inverse Problems 34 (1) (2017) 014004. doi:10.1088/1361-6420/aa9a90.
URL https://doi.org/10.1088%2F1361-6420%2Faa9a90

[48] T. Qin, Z. Chen, J. D. Jakeman, D. Xiu, Deep learning of parameterized equations with applications to uncertainty quantification, International Journal for Uncertainty Quantification 11 (2) (2021) 63–82.
https://epubs.siam.org/doi/pdf/10.1137/1.9781611971200, doi:10.1137/1.9781611971200.
URL https://epubs.siam.org/doi/abs/10.1137/1.9781611971200

[49] R. Koenker, Quantile Regression, Econometric Society Monographs, Cambridge University Press, 2005.

[50] Y. Cui, J. Pang, Modern Nonconvex Nondifferentiable Optimization, Mos-siam series on optimization, Society for Industrial and Applied Mathematics, 2021.
URL https://books.google.com/books?id=s958zgEACAAJ

[51] R. DeVore, G. Lorentz, Constructive Approximation, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 1993.
URL https://books.google.com/books?id=cDqNW6k7_ZwC

[52] J. Lee, S. Leyffer, Mixed Integer Nonlinear Programming, The IMA Volumes in Mathematics and its Applications, Springer New York, 2011.
URL https://books.google.com/books?id=ja0R-L-YCxIC

[53] S. Conti, M. Rumpf, R. Schultz, S. Tölkes, Stochastic dominance constraints in elastic shape optimization, SIAM Journal on Control and Optimization 56 (4) (2018) 3021–3034. doi:10.1137/16M108313X.

[54] J. D. Jakeman, *Pyapprox: Approximation and probabilistic analysis of data.* https://sandialabs.github.io/pyapprox/index.html, 2019.

[55] Z. Liu, A. Narayan, On the computation of recurrence coefficients for univariate orthogonal polynomials (2021). arXiv:2101.11963.

[56] M. Norton, V. Khokhlov, S. Uryasev, Calculating CVaR and bPOE for common probability distributions with application to portfolio optimization and density estimation, Annals of Operations Research (2019) 1–35.

[57] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[58] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4):877–900, 1999.

[59] V. H. Hoang, C. Schwab, Regularity and generalized polynomial chaos approximation of parametric and random second-order hyperbolic partial differential equations, Analysis and Applications 10 (03) (2012) 295–326. https://doi.org/10.1142/S0219530512500145, doi:10.1142/S0219530512500145.
URL https://doi.org/10.1142/S0219530512500145

[60] R. A. Todor, C. Schwab, Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients, IMA Journal of Numerical Analysis 27 (2) (2007) 232–261. https://academic.oup.com/imajna/article-pdf/27/2/232/1980506/drl025.pdf, doi:10.1093/imanum/drl025.
URL https://doi.org/10.1093/imanum/drl025

[61] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, Comptes Rendus Mécanique 336 (6) (2008) 518–523. doi:10.1016/j.crme.2008.02.013.

[62] R. Schöbi, B. Sudret, S. Marelli, Rare event estimation using polynomial-chaos kriging, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 3 (2) (2017) D4016002. https://ascelibrary.org/doi/pdf/10.1061/AJRUA6.0000870, doi:10.1061/AJRUA6.0000870.
URL https://ascelibrary.org/doi/abs/10.1061/AJRUA6.0000870

[63] Ernst, O. G., Mugler, A., Starkloff, H.-J., Ullmann, E., On the convergence of generalized polynomial chaos expansions, ESAIM: M2AN 46 (2) (2012) 317–339. doi:10.1051/m2an/2011045.
URL https://doi.org/10.1051/m2an/2011045

[64] M. Vohra, P. Nath, S. Mahadevan, Y.-T. Tina Lee, Fast surrogate modeling using dimensionality reduction in model inputs and field output: Application to additive manufacturing, Reliability Engineering & System Safety 201 (2020) 106986. doi:https://doi.org/10.1016/j.ress.2020.106986.
URL https://www.sciencedirect.com/science/article/pii/S0951832020304877

[65] M. M. Zuniga, A. Murangira, T. Perdrizet, Structural reliability assessment through surrogate based importance sampling with dimension reduction, Reliability Engineering & System Safety 207 (2021) 107289. doi:https://doi.org/10.1016/j.ress.2020.107289.
URL https://www.sciencedirect.com/science/article/pii/S0951832020307857

[66] F. A., A. Sobester, A. Keane, Engineering Design via Surrogate Modelling: A Practical Guide, John Wiley & Sons, Ltd, 2008. doi:10.1002/9780470770801.

[67] P. G. Constantine, E. Dow, Q. Wang, Active subspace methods in theory and practice: Applications to kriging surfaces, SIAM Journal on Scientific Computing 36 (4) (2014) A1500–A1524. https://doi.org/10.1137/130916138, doi:10.1137/130916138.
URL https://doi.org/10.1137/130916138

[68] M. Heinkenschloss, B. Kramer, T. Takhtaganov, Adaptive reduced-order model construction for conditional value-at-risk estimation, SIAM/ASA Journal on Uncertainty Quantification 8 (2) (2020) 668–692. https://doi.org/10.1137/19M1257433, doi:10.1137/19M1257433.
URL https://doi.org/10.1137/19M1257433