SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 6, No. 4, pp. 1395–1423

Conditional-Value-at-Risk Estimation via Reduced-Order Models*

Matthias Heinkenschloss†, Boris Kramer‡, Timur Takhtaganov§, and Karen Willcox¶

Abstract. This paper proposes and analyzes two reduced-order model (ROM) based approaches for the efficient and accurate evaluation of the Conditional-Value-at-Risk (CVaR) of quantities of interest (QoI) in engineering systems with uncertain parameters. CVaR is used to model objective or constraint functions in risk-averse engineering design and optimization applications under uncertainty. Evaluating the CVaR of the QoI requires sampling in the tail of the QoI distribution and typically requires many solutions of an expensive full-order model of the engineering system. Our ROM approaches substantially reduce this computational expense. Both ROM-based approaches use Monte Carlo (MC) sampling. The first approach replaces the computationally expensive full-order model by an inexpensive ROM. The resulting CVaR estimation error is proportional to the ROM error in the so-called risk region, a small region in the space of uncertain system inputs. The second approach uses a combination of full-order model and ROM evaluations via importance sampling and is effective even if the ROM has large errors. In the importance sampling approach, ROM samples are used to estimate the risk region and to construct a biasing distribution. A few full-order model samples are then drawn from this biasing distribution. Asymptotically, as the ROM error goes to zero, the importance sampling estimator reduces the variance by a factor of $1 - \beta \ll 1$, where $\beta \in (0, 1)$ is the quantile level at which CVaR is computed. Numerical experiments on a system of semilinear convection-diffusion-reaction equations illustrate the performance of the approaches.

Key words. reduced-order models, risk measures, conditional-value-at-risk, estimation, sampling

AMS subject classifications. 35R60, 62H12, 65G99, 65Y20

DOI. 10.1137/17M1160069

1. Introduction.

Designing reliable engineering systems requires taking into account the uncertainties associated with system parameters. In risk-averse optimization, so-called risk measures are applied to quantities of interest (QoI) X to form the objective function and constraint functions. This paper proposes rigorous ways to use reduced-order models (ROMs) for the the efficient estimation of the so-called Conditional-Value-at-Risk at level β (CVaR_β), $\beta \in (0, 1)$. CVaR_β builds on the Value-at-Risk at level β (VaR_β), which is the β-quantile of the random variable X. If X represents a loss (or cost or target violation), $\text{VaR}_\beta[X]$

*Received by the editors December 6, 2017; accepted for publication (in revised form) July 30, 2018; published electronically October 16, 2018.
http://www.siam.org/journals/juq/6-4/M116006.html
Funding: This research was supported in part by NSF grant DMS-1522798 and by the DARPA EQUiPS Program, Award UTA15-001068.

†Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005-1892 (heinken@rice.edu).
‡Department of Aeronautics and Astronautics, MIT, Cambridge, MA 02139 (bokramer@mit.edu, kwillcox@mit.edu).
§Center for Computational Sciences and Engineering, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 (timur.takhtaganov@gmail.com). At the time of writing, the author was a Ph.D. candidate at Rice University.
¶Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712 (kwillcox@ices.utexas.edu).

is the smallest value so that losses over this value occur with probability $1-\beta$. VaR_β is not convex and it does not account for size of losses. $\mathrm{CVaR}_\beta[X]$ removes those drawbacks. $\mathrm{CVaR}_\beta[X]$ is the average of the losses above VaR_β scaled by $1/(1-\beta)$ and is convex. Originally used as a risk measure in financial applications, CVaR_β is now also more frequently used in engineering applications; see, e.g., the survey paper by Rockafellar and Royset [14]. Estimating CVaR_β (and many other risk measures) requires sampling from the tail of the distribution of the QoI and is computationally expensive. While the distribution of the uncertain system parameters is known, the resulting QoI is a random variable that depends on the solution of the system. Therefore, the distribution of the QoI is not explicitly known, and naive sampling from the tail of this distribution to estimate CVaR_β requires many expensive system simulations. Each system simulation requires the expensive solution of a full-order model (FOM). In our application, the system is modeled by a system of partial differential equations (PDEs), and the FOM is a high-fidelity discretization of the PDEs.

This paper proposes and analyzes two ROM based approaches for the efficient and accurate estimation of CVaR_β. In our first approach we replace the FOM by a ROM, and we analyze the resulting error in the CVaR_β estimate. This error analysis depends on so-called risk regions, small regions in parameter space that contain the parameters that lead to "risky" QoI values. Specifically, we show that the error in the CVaR_β estimate due to ROMs is upper bounded by the ROM error in so-called ϵ-risk regions associated with the ROMs. These ϵ-risk regions are derived from ROM information, contain the risk region of the FOM, and as the ROM error goes to zero converges to the risk region of the FOM. This first approach relies on an error bound between the ROM and the FOM, and the error needs to be sufficiently small for good CVaR_β estimates.

Our second approach uses an importance sampling framework. Here, the ROM error bound is only used to obtain estimates of the risk region. The proposed importance sampling framework can be effective even if the ROM error is large. ROMs are used to construct a biasing distribution that biases samples towards the risk region. We then estimate CVaR_β from FOM samples drawn from this biasing distribution. The ROM approximation error only impacts the quality of the biasing distribution and therefore, via its variance, the sampling error. The smaller this variance, the fewer FOM samples are needed to generate a CVaR_β estimate at a desired tolerance. We show that asymptotically as the ROM error goes to zero, the importance sampling estimator reduces the variance by a factor of $1-\beta \ll 1$, where $\beta \in (0,1)$ is the quantile level at which CVaR is computed.

There are existing approaches for using ROMs in the estimation of CVaR_β and other risk measures. Variance reduction strategies (including importance sampling) for quantile (Value-at-Risk) estimation using reduced models are considered by Cannamela, Garnier, and Iooss [2]. Specifically, importance-sampling estimates are constructed using a biasing density from a parametrized family of densities, and the parameters are informed through a reduced model. In this work, we consider CVaR_β (superquantiles), which are based on VaR_β (quantiles), and we analyze the impact of the ROM error on the CVaR_β estimate. Harajli, Rockafellar, and Royset [5] compute buffered failure probabilities (which are related to CVaR) via importance sampling. However, an analytic biasing density is chosen which depends on the specific test problem in [5]. We suggest a principled, general approach to obtain a biasing density via reduced-order models obtained from the original governing equations. Importance sampling

via reduced models has been successfully applied in the context of computing failure probabilities by Li, Li, and Xiu [8] and Peherstorfer et al. [10, 11, 12]. However, these approaches do not directly translate to importance sampling for CVaR_β estimation, the focus of our work. The review by Hong, Hu, and Liu [6, sects. 2.1, 2.2] states asymptotic properties of CVaR_β estimation via importance sampling given a biasing density, but does not address how one can compute a biasing density. Among other contributions, our paper proposes a construction of a biasing density using ROMs. Zou, Kouri, and Aquino [22] estimate CVaR_β by constructing a Voronoi tessellation of the parameter space and using localized reduced-basis surrogate models. While their ROM construction is different from what we use in our example, it could, in principle, be used as well, and their overall approach is comparable to our first approach. The main difference between the work [22] and our work is that the ROM error in the entire parameter region is needed in [22], whereas our error estimate depends only on the ROM error in the risk region. In addition, we introduce and analyze an importance sampling approach that uses ROMs to construct the biasing distributions. Proper orthogonal decomposition based ROMs have recently been used in [21] to minimize CVaR_β for an aircraft noise problem modeled by the Helmholtz equation. Their overall approach is comparable to our first approach, but they do not analyze the impact of ROMs on the CVaR_β estimation error. The design of an ultrahigh-speed hydrofoil by using CVaR_β optimization is considered by Royset et al. [17]. They propose building surrogates of the CVaR of their QoI and model these surrogates as random variables "due to unknown error in the surrogate relative to the actual value" of the CVaR of their QoI. This randomness in the CVaR surrogate is then incorporated into the design process by applying CVaR again, but with a different quantile level to the surrogate. Ultimately, they use a surrogate for the QoI that combines high-fidelity and low-fidelity QoI evalutions into a polynomial fit model. Our work does not require additional stochastic treatment of model error, and focuses on the efficient and accurate sampling of CVaR using ROMs of the QoI that satisfy the original governing equations.

This paper is structured as follows. Section 2 provides background material and notation for CVaR_β computation. In section 3 we derive error bounds for CVaR_β estimation with ROMs and give results on confidence intervals for CVaR_β. Section 4 presents our results on CVaR_β estimation with importance sampling. Our two ROM approaches are illustrated on a system of semilinear convection-diffusion-reaction PDEs in section 5. A review and discussion is given in section 6.

2. Problem formulation and background.

This section specifies our problem set-up and reviews the results on CVaR_β needed for our application of ROMs. Section 2.1 introduces the basic problem set-up and notation, followed by a brief introduction to CVaR_β and some useful properties in section 2.2. Sample estimates of VaR_β and CVaR_β together with a complete algorithm are given in section 2.3.

2.1. The state equation and quantity of interest.

We consider systems modeled by equations of the form

$$F(y, \boldsymbol{\xi}) = 0, \tag{2.1}$$

where $\boldsymbol{\xi}$ is a vector of random variables (continuous or discrete) with values in $\Xi \subset \mathbb{R}^M$ and with probability density function (p.d.f.) ρ, and where y denotes the state of the system.

Equation (2.1) is referred to as the state equation. Often the system is modeled by a system of PDEs in which case the state equation (2.1) is a high-fidelity discretization of the PDEs. We assume that for every realization $\xi \in \Xi$ there exists a unique solution $y = y(\xi) \in \mathbb{R}^N$ of (2.1). For discretized PDEs, N is typically large.

We are interested in a QoI $s : \mathbb{R}^N \mapsto \mathbb{R}$, and we assume that

$$\xi \mapsto s(y(\xi))$$

is both in $L^1_\rho(\Xi)$ and in $L^2_\rho(\Xi)$. For notational convenience, we set the QoI to be the random variable

$$X = s(y(\cdot)). \tag{2.2}$$

Following the original setting of financial applications, it is helpful to think of high values of X as risky.

The expected value and variance of the random variable X are given by

$$\mathbb{E}_\rho[X] = \int_\Xi X(\xi)\rho(\xi)d\xi \quad \text{and} \quad \mathbb{V}_\rho[X] = \mathbb{E}_\rho\left[(X(\cdot) - \mathbb{E}_\rho[X])^2\right],$$

respectively. The subscript ρ is used to indicate which density is used in the integration.

2.2. Conditional-Value-at-Risk (CVaR). The CVaR_β is based on the Value-at-Risk (VaR_β). For a given level $\beta \in (0,1)$ the $\mathrm{VaR}_\beta[X]$ is the β-quantile of the random variable X,

$$\mathrm{VaR}_\beta[X] = \min_{t\in\mathbb{R}}\{\Pr[X \le t] \ge \beta\}. \tag{2.3}$$

Here

$$\Pr[X \le t] = \int_\Xi \mathbb{I}\{X(\xi) \le t\}\rho(\xi)d\xi,$$

the probability that X is less or equal than t and $\mathbb{I}$ is the indicator function. Different equivalent definitions of CVaR_β exist. The following definition is due to Rockafellar and Uryasev [15, 16]. The CVaR_β at level $\beta \in (0,1)$ is

$$\mathrm{CVaR}_\beta[X] = \min_{t\in\mathbb{R}}\left\{t + \frac{1}{1-\beta}\,\mathbb{E}_\rho[(X-t)_+]\right\}, \tag{2.4}$$

where $(\cdot)_+ = \max\{\cdot, 0\}$. The minimum of (2.4) is attained on the interval $[t^*, t^{**}]$ with

$$t^* = \mathrm{VaR}_\beta[X] \quad \text{and} \quad t^{**} = \sup\{t : \Pr[X \le t] \le \beta\}.$$

Inserting $t^* = \mathrm{VaR}_\beta[X]$ into (2.4), one obtains

$$\mathrm{CVaR}_\beta[X] = \mathrm{VaR}_\beta[X] + \frac{1}{1-\beta}\,\mathbb{E}_\rho\left[(X - \mathrm{VaR}_\beta[X])_+\right]. \tag{2.5}$$

If the cumulative distribution function (c.d.f.) $H_X(x) = \Pr[X \le x]$ is continuous at $x = \mathrm{VaR}_\beta[X]$, then $\Pr\big[X = \mathrm{VaR}_\beta[X]\big] = 0$, $t^* = t^{**} = \mathrm{VaR}_\beta[X]$, and (2.5) can be simplified to

$$\mathrm{CVaR}_\beta[X] = \frac{1}{1-\beta}\,\mathbb{E}_\rho\left[X \cdot \mathbb{I}\{X \ge \mathrm{VaR}_\beta[X]\}\right]. \tag{2.6}$$

More generally, the coherent risk measure properties of CVaR_β and the biconjugate representation of coherent risk measures (see, e.g., [18], [20, sect. 6.3], and [7]) give the representation

$$\mathrm{CVaR}_\beta[X] = \sup_{\vartheta \in \mathcal{A}} \mathbb{E}_\rho\left[\vartheta X\right], \tag{2.7a}$$

where

$$\mathcal{A} = \left\{\vartheta \in L_\rho^\infty(\Xi) \;:\; 0 \le \vartheta(\xi) \le (1-\beta)^{-1} \text{ a.e. } \Xi, \text{ and } \mathbb{E}_\rho\left[\vartheta\right] = 1\right\}. \tag{2.7b}$$

One can show [20, Example 6.16] that the supremum is attained at any ϑ_X^* that satisfies

$$\vartheta_X^*(\xi) \begin{cases} = 0 & \text{if } X(\xi) < \mathrm{VaR}_\beta[X], \\ \in [0, (1-\beta)^{-1}] & \text{if } X(\xi) = \mathrm{VaR}_\beta[X], \\ = (1-\beta)^{-1} & \text{if } X(\xi) > \mathrm{VaR}_\beta[X], \end{cases} \tag{2.8}$$

for almost all $\xi \in \Xi$, and $\mathbb{E}_\rho\left[\vartheta_X^*\right] = 1$. If $\Pr\left[X = \mathrm{VaR}_\beta[X]\right] = 0$, ϑ_X^* is unique and (2.8), (2.7) reduce to (2.6).

The representations (2.8), (2.7), and (2.6) show that $\mathrm{CVaR}_\beta[X]$ depends only on the values of X that lie in the upper tail of the c.d.f. Therefore, the values of the parameter vector ξ that correspond to these high values of X can be considered risky. This motivates the following definition.

Definition 2.1. *The risk region corresponding to $CVaR_\beta[X]$ is given by*

$$\mathbb{G}_\beta[X] := \{\xi \mid X(\xi) \ge VaR_\beta[X]\} \subset \Xi$$

and the corresponding indicator function of the risk region $\mathbb{G}_\beta[X]$ is

$$\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi) := \mathbb{I}\{X(\xi) \ge VaR_\beta[X]\}. \tag{2.9}$$

Note that

$$\begin{aligned} \Pr\left[\mathbb{G}_\beta[X]\right] &= \Pr\left[\{X > \mathrm{VaR}_\beta[X]\}\right] + \Pr\left[\{X = \mathrm{VaR}_\beta[X]\}\right] \\ &= (1-\beta) + \Pr\left[\{X = \mathrm{VaR}_\beta[X]\}\right] \end{aligned} \tag{2.10}$$

by the definition (2.3) of $\mathrm{VaR}_\beta[X]$.

2.3. Sampling-based estimation of VaR_β and CVaR_β. In practice, $\mathrm{CVaR}_\beta[X]$ is estimated through sampling. Throughout this paper we use $\widehat{\cdot}$ to denote a sampling-based estimate. For example, $\widehat{\mathrm{CVaR}}_\beta[X]$ and $\widehat{\mathrm{VaR}}_\beta[X]$ denote sampling-based estimates of $\mathrm{CVaR}_\beta[X]$ and $\mathrm{VaR}_\beta[X]$, respectively. The procedure for computing estimates of $\mathrm{VaR}_\beta[X]$ and $\mathrm{CVaR}_\beta[X]$ is described in Algorithm 2.1.[1] We will use Algorithm 2.1 for standard MC sampling, in which case $p^{(n)} = 1/n$, or with importance sampling (see section 4), in which case $p^{(n)} = w(\xi^{(n)})/n$ with w being the weight function in the importance sampling. The second term on the right-hand side in (2.12) is nonzero for the case $\sum_{j=1}^{k_\beta - 1} p^{(j)} \ne 1 - \beta$ and is based on the idea of splitting the probability atom at $\mathrm{VaR}_\beta[X]$ (see [16]).

[1] Rockafellar and Uryasev [16] assume that samples are ordered $X(\xi^{(1)}) < X(\xi^{(2)}) < \cdots < X(\xi^{(n)})$ with corresponding probabilities $p^{(1)}, \ldots, p^{(n)}$. Since $\mathrm{VaR}_\beta[X]$ and $\mathrm{CVaR}_\beta[X]$ depend only on (a few) samples with large values, we prefer the ordering of (2.11).

Algorithm 2.1. Sampling-based estimation of VaR_β and CVaR_β.

Input: Parameter samples $\xi^{(1)}, \ldots, \xi^{(n)}$ with probabilities $p^{(1)}, \ldots, p^{(n)}$, risk level $\beta \in (0,1)$, and random variable $X(\xi)$.

Output: Estimate $\widehat{\mathrm{VaR}}_\beta[X]$ and $\widehat{\mathrm{CVaR}}_\beta[X]$.

1: Evaluate X at the parameter samples: $X(\xi^{(1)}), \ldots, X(\xi^{(n)})$.

2: Sort values of X in descending order and relabel the samples so that

$$X(\xi^{(1)}) > X(\xi^{(2)}) > \cdots > X(\xi^{(n)}), \tag{2.11}$$

and reorder the probabilities accordingly (so that $p^{(j)}$ corresponds to $\xi^{(j)}$).

3: Compute an index k_β such that

$$\sum_{j=1}^{k_\beta - 1} p^{(j)} \le 1 - \beta < \sum_{j=1}^{k_\beta} p^{(j)}.$$

4: Set

$$\widehat{\mathrm{VaR}}_\beta[X] = X(\xi^{(k_\beta)})$$

and

$$\widehat{\mathrm{CVaR}}_\beta[X] = \frac{1}{1-\beta} \sum_{j=1}^{k_\beta - 1} p^{(j)} X(\xi^{(j)}) + \frac{1}{1-\beta} \left(1 - \beta - \sum_{j=1}^{k_\beta - 1} p^{(j)} \right) \widehat{\mathrm{VaR}}_\beta[X]. \tag{2.12}$$

Asymptotic convergence properties of the estimators $\widehat{\mathrm{VaR}}_\beta[X]$ and $\widehat{\mathrm{CVaR}}_\beta[X]$ are given in the review by Hong, Hu, and Liu [6, sects. 2.1, 2.2] and we state some of these later in Lemma 4.3 when we discuss importance sampling.

3. CVaR_β estimation with reduced-order models.

Recall that evaluating the QoI $X(\xi)$ at a given parameter ξ requires solving an expensive FOM. To devise our computationally efficient framework, we assume the availability of an inexpensive ROM approximation of the parameter to QoI map, denoted as $X_r(\xi)$. For the purposes of this section, the details of how $X_r(\xi)$ is computed are not important. Later, in section 5, we give an example of how $X_r(\xi)$ is computed via state reduction for the convection-diffusion-reaction model problem.

Our first proposed approach, described in this section, approximates $\mathrm{CVaR}_\beta[X]$ by $\mathrm{CVaR}_\beta[X_r]$. Since ROM samples $X_r(\xi)$ are relatively inexpensive to compute, the computation of $\mathrm{CVaR}_\beta[X_r]$ is relatively inexpensive. Section 3.1 presents estimates of the error between $\mathrm{CVaR}_\beta[X]$ and $\mathrm{CVaR}_\beta[X_r]$. Section 3.2 then states the algorithm for practical computation of the ROM errors for general nonlinear systems.

3.1. Error analysis for CVaR_β estimation with ROM.

For the purpose of the following derivation, assume the availability of a bound for the error between the original QoI X and

its approximation X_r,

$$|X(\xi) - X_r(\xi)| \le \epsilon_r(\xi) \quad \text{for} \quad \xi \in \Xi. \tag{3.1}$$

We will relax this assumption somewhat at the end of this section.

Define the maximum error in Ξ as

$$\epsilon_r^{\max} := \sup_{\xi \in \Xi} \epsilon_r(\xi) \tag{3.2}$$

so that $|X(\xi) - X_r(\xi)| \le \epsilon_r^{\max}$ for all $\xi \in \Xi$. Zou, Kouri, and Aquino [22] use this maximum error and the biconjugate representation of CVaR_β to derive an error estimate for CVaR_β. First, using (2.7) gives

$$\begin{aligned}\mathrm{CVaR}_\beta[X] &= \sup_{\vartheta \in \mathcal{A}} \mathbb{E}[\vartheta X] \le \sup_{\vartheta \in \mathcal{A}} \mathbb{E}[\vartheta X_r] + \sup_{\vartheta \in \mathcal{A}} \mathbb{E}\left[\vartheta |X - X_r|\right] \\ &= \mathrm{CVaR}_\beta[X_r] + \sup_{\vartheta \in \mathcal{A}} \mathbb{E}\left[\vartheta |X - X_r|\right] \le \mathrm{CVaR}_\beta[X_r] + \epsilon_r^{\max},\end{aligned}$$

where the final inequality follows from the fact that elements $\vartheta \in \mathcal{A}$ are probability densities. Similarly, $\mathrm{CVaR}_\beta[X_r] \le \mathrm{CVaR}_\beta[X] + \epsilon_r^{\max}$. Thus,

$$\left|\mathrm{CVaR}_\beta[X] - \mathrm{CVaR}_\beta[X_r]\right| \le \sup_{\vartheta \in \mathcal{A}} \mathbb{E}\left[\vartheta |X - X_r|\right] \le \epsilon_r^{\max}. \tag{3.3}$$

Although $\mathrm{CVaR}_\beta[X]$ and $\mathrm{CVaR}_\beta[X_r]$ depend only on X and X_r in their respective risk regions, the error bound (3.3) depends on the maximum error in Ξ. In particular, if the distributions of X and X_r have identical tails but differ elsewhere in Ξ, the actual error $|\mathrm{CVaR}_\beta[X] - \mathrm{CVaR}_\beta[X_r]|$ will be zero, whereas the error bound in (3.3) will still be $\epsilon_r^{\max}$. It turns out that only the error between X and X_r in the so-called ϵ-risk region, defined next, is important.

Definition 3.1. *The ϵ-risk region corresponding to $\mathrm{CVaR}_\beta[X]$ is given by*

$$\mathbb{G}_\beta^\epsilon[X_r] := \{\xi \,:\, X_r(\xi) + \epsilon_r(\xi) \ge \mathrm{VaR}_\beta[X_r - \epsilon_r]\}, \tag{3.4}$$

and the corresponding indicator function is $\mathbb{I}_{\mathbb{G}_\beta^\epsilon[X_r]}(\xi)$.

Lemma 3.2. *It holds that*

$$\mathbb{G}_\beta[X] \subseteq \mathbb{G}_\beta^\epsilon[X_r] \quad \text{and} \quad \mathbb{G}_\beta[X_r] \subseteq \mathbb{G}_\beta^\epsilon[X_r] \tag{3.5}$$

and

$$\epsilon_r^{G,low} := \sup_{\xi \in \mathbb{G}_\beta[X_r]} \epsilon_r(\xi) \le \sup_{\xi \in \mathbb{G}_\beta[X] \cup \mathbb{G}_\beta[X_r]} \epsilon_r(\xi) \le \sup_{\xi \in \mathbb{G}_\beta^\epsilon[X_r]} \epsilon_r(\xi) =: \epsilon_r^G. \tag{3.6}$$

Proof. Consider the random variables $X_r - \epsilon_r$ and $X_r + \epsilon_r$. Obviously,

$$X_r(\xi) - \epsilon_r(\xi) \le X(\xi) \le X_r(\xi) + \epsilon_r(\xi) \quad \forall \xi \in \Xi,$$

and, by monotonicity of VaR_β (see, e.g., [13]),

$$\mathrm{VaR}_\beta[X_r - \epsilon_r] \le \mathrm{VaR}_\beta[X] \le \mathrm{VaR}_\beta[X_r + \epsilon_r].$$

In particular, for $\xi \in \mathbb{G}_\beta[X]$ we have

$$\mathrm{VaR}_\beta[X_r - \epsilon_r] \le \mathrm{VaR}_\beta[X] \le X(\xi) \le X_r(\xi) + \epsilon_r(\xi),$$

which implies that $\mathbb{G}_\beta[X] \subseteq \mathbb{G}^\epsilon_\beta[X_r]$. Similarly, for $\xi \in \mathbb{G}_\beta[X_r]$ we have

$$\mathrm{VaR}_\beta[X_r - \epsilon_r] \le \mathrm{VaR}_\beta[X_r] \le X_r(\xi) \le X_r(\xi) + \epsilon_r(\xi),$$

i.e., that $\mathbb{G}_\beta[X_r] \subseteq \mathbb{G}^\epsilon_\beta[X_r]$. Inequality (3.6) is an immediate consequence of (3.5). ∎

We can now make the opening statement of this section more precise. In particular, we do not need the error function $\epsilon_r(\xi)$ in all of Ξ from (3.1), but only the error ϵ_r^G in the ϵ-risk region $\mathbb{G}^\epsilon_\beta[X_r]$ from (3.6).

Theorem 3.3. *The error between $CVaR_\beta$ of the full-order model X and $CVaR_\beta$ of the reduced-order model X_r is bounded as*

$$\begin{aligned} &\big|CVaR_\beta[X] - CVaR_\beta[X_r]\big| \\ &\le \left(1 + \frac{\sup\big\{\Pr\big[\{X = VaR_\beta[X]\}\big],\ \Pr\big[\{X_r = VaR_\beta[X_r]\}\big]\big\}}{1-\beta}\right) \epsilon_r^G \\ &\le \left(1 + \frac{1}{1-\beta}\right) \epsilon_r^G. \end{aligned} \tag{3.7}$$

If X and X_r have c.d.f.'s that are continuous at $VaR_\beta[X]$ and at $VaR_\beta[X_r]$, respectively, then

$$\big|CVaR_\beta[X] - CVaR_\beta[X_r]\big| \le \epsilon_r^G. \tag{3.8}$$

Proof. Let ϑ_X^* be given by (2.8) with $\mathrm{CVaR}_\beta[X] = \sup_{\vartheta\in\mathcal{A}} \mathbb{E}[\vartheta X] = \mathbb{E}[\vartheta_X^* X]$. Then using the definition of ϑ_X^*, we see that $\mathbb{E}[\vartheta_X^* X_r] \le \sup_{\vartheta\in\mathcal{A}} \mathbb{E}[\vartheta X_r] = \mathrm{CVaR}_\beta[X_r]$, and (2.10) gives

$$\begin{aligned} \mathrm{CVaR}_\beta[X] &= \mathbb{E}[\vartheta_X^* X] = \mathbb{E}[\vartheta_X^* X_r] + \mathbb{E}[\vartheta_X^*(X - X_r)] \\ &\le \mathrm{CVaR}_\beta[X_r] + \frac{1}{1-\beta}\int_{\mathbb{G}_\beta[X]} |X - X_r|\rho(\xi)d\xi \\ &\le \mathrm{CVaR}_\beta[X_r] + \frac{1-\beta+\Pr\big[\{X = \mathrm{VaR}_\beta[X]\}\big]}{1-\beta} \sup_{\xi\in\mathbb{G}_\beta[X]} \epsilon_r(\xi). \end{aligned}$$

Similarly,

$$\mathrm{CVaR}_\beta[X_r] \le \mathrm{CVaR}_\beta[X] + \frac{1-\beta+\Pr\big[\{X_r = \mathrm{VaR}_\beta[X_r]\}\big]}{1-\beta} \sup_{\xi\in\mathbb{G}_\beta[X_r]} \epsilon_r(\xi).$$

It follows that

$$\begin{aligned}&\big|\mathrm{CVaR}_\beta[X] - \mathrm{CVaR}_\beta[X_r]\big| \\ &\le \left(1 + \frac{\max\big\{\Pr\big[\{X = \mathrm{VaR}_\beta[X]\}\big],\ \Pr\big[\{X_r = \mathrm{VaR}_\beta[X_r]\}\big]\big\}}{1-\beta}\right) \sup_{\xi \in \mathbb{G}_\beta[X] \cup \mathbb{G}_\beta[X_r]} \epsilon_r(\xi).\end{aligned} \tag{3.9}$$

The first term on the right-hand side of (3.9) is bounded by $1 + 1/(1-\beta)$. If X and X_r have c.d.f's that are continuous at $\mathrm{VaR}_\beta[X]$ and $\mathrm{VaR}_\beta[X_r]$, respectively, then $\Pr\big[\{X = \mathrm{VaR}_\beta[X]\}\big] = \Pr\big[\{X_r = \mathrm{VaR}_\beta[X_r]\}\big] = 0$, and the first term becomes one. The second term on the right-hand side of (3.9) is bounded by ϵ_r^G (see (3.6)). ∎

Algorithm 3.1. Sampling-based estimation of CVaR_β and its errors with ROM.

Input: Parameter samples $\xi^{(1)}, \ldots, \xi^{(n)}$ with probabilities $p^{(1)}, \ldots, p^{(n)}$, risk level β, random variable $X_r(\xi)$, and its error function $\epsilon_r(\xi)$.

Output: Estimate $\widehat{\mathrm{CVaR}}_\beta[X_r]$ and error estimates $\widehat{\epsilon}_r^{\max}$, $\widehat{\epsilon}_r^{G,\mathrm{low}}$, and $\widehat{\epsilon}_r^{G}$.

1: Evaluate X_r and ϵ_r at the parameter samples: $X_r(\xi^{(1)}), \ldots, X_r(\xi^{(n)})$, $\epsilon_r(\xi^{(1)}), \ldots, \epsilon_r(\xi^{(n)})$.

2: Apply steps 2, 3, and 4 of Algorithm 2.1 with X replaced by X_r to obtain the index k_β and $\widehat{\mathrm{CVaR}}_\beta[X_r]$.

3: Reorder $\epsilon_r(\xi^{(1)}), \ldots, \epsilon_r(\xi^{(n)})$ to match the order of $X_r(\xi^{(1)}), \ldots, X_r(\xi^{(n)})$ from the previous step.

4: Approximate $\epsilon_r^{\max}$ and $\epsilon_r^{G,\mathrm{low}}$:

$$\widehat{\epsilon_r}^{\max} = \max_{1 \le j \le n} \epsilon_r(\xi^{(j)}), \qquad \widehat{\epsilon_r}^{G,\mathrm{low}} = \max_{1 \le j \le k_\beta} \epsilon_r(\xi^{(j)}).$$

5: Estimate $\widehat{\mathrm{VaR}}_\beta[X_r - \epsilon_r]$ by applying steps 2, 3, and 4 of Algorithm 2.1 with X replaced by $X_r - \epsilon_r$.

6: Approximate the ϵ-risk region (3.4) by a discrete set:

$$\widehat{\mathbb{G}_\beta^\epsilon}[X_r] := \{\xi^{(j)} \,:\, X_r(\xi^{(j)}) + \epsilon_r(\xi^{(j)}) \ge \widehat{\mathrm{VaR}}_\beta[X_r - \epsilon_r]\}.$$

7: Compute

$$\widehat{\epsilon}_r^{G} = \max_{\xi^{(j)} \in \widehat{\mathbb{G}_\beta^\epsilon}[X_r]} \epsilon_r(\xi^{(j)}).$$

Remark 3.4. The bound (3.7) uses the crude bounds $\Pr\big[\{X = \mathrm{VaR}_\beta[X]\}\big] \le 1$ and $\Pr\big[\{X_r = \mathrm{VaR}_\beta[X_r]\}\big] \le 1$. This can be improved, e.g., as follows.

The $\mathrm{VaR}_\beta[X]$ is monotonically increasing in β. For any $\delta \in (0, \beta)$ such that $\mathrm{VaR}_{\beta-\delta}[X] < \mathrm{VaR}_\beta[X] < \mathrm{VaR}_{\beta+\delta}[X]$ it holds that

$$\begin{aligned}\Pr\big[\{X = \mathrm{VaR}_\beta[X]\}\big] &\le \Pr\big[\{X > \mathrm{VaR}_{\beta-\delta}[X]\} \cap \{X \le \mathrm{VaR}_{\beta+\delta}[X]\}\big] \\ &= \Pr\big[\{X > \mathrm{VaR}_{\beta-\delta}[X]\}\big] - \Pr\big[\{X > \mathrm{VaR}_{\beta+\delta}[X]\}\big] \\ &= 1 - (\beta - \delta) - (1 - (\beta + \delta)) = 2\delta.\end{aligned}$$

Note that if the c.d.f. of X is continuous at $\mathrm{VaR}_\beta[X]$, then $\mathrm{VaR}_{\beta-\delta}[X] < \mathrm{VaR}_\beta[X] < \mathrm{VaR}_{\beta+\delta}[X]$ for any $\delta \in (0, \beta)$. The term $\Pr\big[\{X_r = \mathrm{VaR}_\beta[X_r]\}\big]$ can be bounded the same way.

3.2. Practical estimation of errors. In practice the ϵ-risk region $\mathbb{G}^\epsilon_\beta[X_r]$ and the errors $\widehat{\epsilon}^{\max}_r$, $\widehat{\epsilon}^{G,\mathrm{low}}_r$, and $\widehat{\epsilon}^{G}_r$ are estimated using sampling. We evaluate the ROM and the full-order model at n random samples and compute the maximum error $\widehat{\epsilon}^{\max}_r$ first. With the ordered samples of X_r we obtain $\widehat{\mathrm{CVaR}}_\beta[X_r]$, and consequently the error $\widehat{\epsilon}^{G,\mathrm{low}}_r$ in the $\mathrm{CVaR}_\beta[X_r]$ region. Next, the ϵ-risk region $\mathbb{G}^\epsilon_\beta[X_r]$ can be computed, which is needed to compute the error $\widehat{\epsilon}^{G}_r$. The details are given in Algorithm 3.1.

4. Importance sampling for estimation of CVaR_β. In the previous section, we used a ROM to replace the FOM in the CVaR_β estimation and showed that the resulting error in CVaR_β is proportional to the ROM error in the ϵ-risk region. Thus, this approach works well if a ROM error estimate is available and the ROM error is sufficiently small. To relax these conditions, we now consider an importance sampling (IS) approach to compute VaR_β and CVaR_β. The ROM is used to generate a so-called biasing density from which samples are drawn. Given this biasing density, few samples of the expensive FOM are used to estimate VaR_β and CVaR_β. While our analysis of the proposed IS approach assumes availability of ROM error bounds to estimate the risk region, our IS approach can be used with fewer assumptions than for the approach in the previous section. Our analysis shows that the performance of our IS approach improves as the ROM error becomes smaller, but as the numerical results in section 5 show, it is effective even with coarse ROMs.

We begin with a brief introduction of the IS framework in section 4.1. Section 4.2 derives the optimal IS density for CVaR_β estimation. This optimal IS density itself is impractical, since it relies on evaluations of the expensive FOM. Therefore, we propose an IS density in section 4.3 that uses the previously introduced ϵ-risk region and, hence, only uses inexpensive evaluations of the ROM. The details of our implementation of the proposed IS approach are provided in section 4.4.

4.1. Importance sampling framework. IS estimators use samples from a biasing distribution (a distribution that is biased towards a specific event, e.g., the risk region) to estimate statistics of the quantity of interest. The estimator accounts for the increased occurrence of such events by including reweighting to compensate within the sampling estimate. For a general introduction to importance sampling, see, e.g., Owen [9, sect. 9]. Mathematically, IS amounts to changing the density, and the following results are stated to allow such a change.

Recall that ρ is the density of the random variable ξ. Define the support $\mathrm{supp}(\rho) = \{\xi \in \Xi \mid \rho(\xi) > 0\}$. Let φ be another density with $\mathrm{supp}(\rho) \subseteq \mathrm{supp}(\varphi)$. For any integrable function $g : \Xi \to \mathbb{R}$ we have

$$\mathbb{E}_\rho[g] = \int_\Xi g(\xi)\ \rho(\xi)\ d\xi = \int_\Xi g(\xi) w(\xi)\ \varphi(\xi)\ d\xi = \mathbb{E}_\varphi[gw],$$

where $w := \rho/\varphi$ is the so-called likelihood ratio, or IS weight function. The subscript φ in $\mathbb{E}_\varphi$ and $\mathbb{V}_\varphi$ indicates that the integrals in the definition of expectation and variance are computed with the density φ.

To derive the IS method for CVaR_β estimation, we make the following assumption throughout this section.

Assumption 4.1. The c.d.f. $H_X(x) = \Pr[X \le x]$ is continuous at $x = \mathrm{VaR}_\beta[X]$.

Under this assumption, $\Pr\big[X = \mathrm{VaR}_\beta[X]\big] = 0$ and

$$\mathrm{CVaR}_\beta[X] = \frac{1}{1-\beta}\int_\Xi X(\xi)\,\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\rho(\xi)d\xi \tag{4.1}$$

(see (2.6) and (2.9)). We note that the assumptions made in Hong, Hu, and Liu [6] to prove asymptotic properties of CVaR_β imply continuity of the c.d.f. $H_X(x)$ at $x = \mathrm{VaR}_\beta[X]$. We emphasize that while this continuity condition is needed to construct the proposed biasing distribution, our IS procedure can be applied even if this assumption does not hold.

We perform a change of measure from the nominal density ρ to the biasing density φ in (4.1) and account for the change by reweighting to obtain

$$\begin{aligned}\mathrm{CVaR}_\beta[X] &= \frac{1}{1-\beta}\int_\Xi \mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)X(\xi)\rho(\xi)d\xi\\ &= \frac{1}{1-\beta}\int_{\widetilde{\Xi}} \mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)X(\xi)w(\xi)\varphi(\xi)d\xi,\end{aligned}$$

where

$$w(\xi) := \frac{\rho(\xi)}{\varphi(\xi)} \tag{4.2}$$

is the weight and $\widetilde{\Xi}$ is the support of the new density φ, to be defined later. Recall that φ does not need to be positive everywhere; it is sufficient (see, e.g., [9, Chapter 9]) that

$$\varphi(\xi) > 0 \quad \text{for } \xi \in \mathbb{G}_\beta[X].$$

Thus we make the following assumption throughout this section.

Assumption 4.2. The support $\widetilde{\Xi}$ of the biasing density φ satisfies

$$\mathbb{G}_\beta[X] \subset \widetilde{\Xi}. \tag{4.3}$$

The IS estimates $\widehat{\mathrm{VaR}}_\beta^{\mathrm{IS}}[X]$ and $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]$ are again computed by Algorithm 2.1, but now we draw independent samples $\xi^{(1)}, \dots, \xi^{(n)}$ from the biasing distribution φ, evaluate $X(\xi^{(j)})$, $j = 1, \dots, n$, and define probabilities $p^{(j)} = w(\xi^{(j)})/n$, $j = 1, \dots, n$. These are now the inputs into Algorithm 2.1 when IS is used.

To justify our choice of the biasing density φ that we will introduce in section 4.3, and to analyze the asymptotic properties of the resulting estimates $\widehat{\mathrm{VaR}}_\beta^{\mathrm{IS}}[X]$ and $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]$, we need the following result adapted from [6, sect. 2.2]. Recall that $\mathbb{E}_\varphi[\cdot]$ and $\mathbb{V}_\varphi[\cdot]$ denote expected value and variance under the measure $\varphi(\xi)d\xi$. In the following result "$\Rightarrow$" denotes convergence in distribution, and $\mathcal{N}(0,1)$ stands for the standard normal distribution.

Lemma 4.3. *If there exists a $\delta > 0$ and $C > 0$ such that X has a positive and continuously differentiable density $h_X(x)$ for all $x \in (VaR_\beta[X]-\delta, VaR_\beta[X]+\delta)$, and if the weight function satisfies $w(\xi) \le C$ for all $\xi \in \{\xi \,|\, X(\xi) \in (VaR_\beta[X]-\delta, \infty)\}$, then $\widehat{VaR}^{IS}_\beta[X] \to VaR_\beta[X]$ w.p. 1 as $n \to \infty$ and*

$$\sqrt{n}\left(\widehat{VaR}^{IS}_\beta[X] - VaR_\beta[X]\right) \Rightarrow \frac{\left(\mathbb{V}_\varphi[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)w(\cdot)]\right)^{1/2}}{h_X(VaR_\beta[X])}\mathcal{N}(0,1).$$

If, in addition, $\mathbb{E}_\varphi[(X(\cdot) - VaR_\beta[X])_+^2\, w^2(\cdot)] < \infty$, then $\widehat{CVaR}^{IS}_\beta[X] \to CVaR_\beta[X]$ w.p. 1 as $n \to \infty$ and

$$\sqrt{n}\left(\widehat{CVaR}^{IS}_\beta[X] - CVaR_\beta[X]\right) \Rightarrow \frac{\left(\mathbb{V}_\varphi[(X(\cdot) - VaR_\beta[X])_+\, w(\cdot)]\right)^{1/2}}{1-\beta}\mathcal{N}(0,1). \tag{4.4}$$

Remark 4.4. Asymptotic results for the standard Monte Carlo estimates, denoted by $\widehat{\mathrm{VaR}}^{\mathrm{MC}}_\beta[X]$ and $\widehat{\mathrm{CVaR}}^{\mathrm{MC}}_\beta[X]$ and computed by Algorithm 2.1 with independent samples $\xi^{(1)}, \ldots, \xi^{(n)}$ from the nominal distribution ρ and equal probabilities $p^{(j)} = 1/n$, $j = 1, \ldots, n$, are a special case of Lemma 4.3 with $w \equiv 1$.

The goal of IS is to compute a biasing density φ such that the variance of the estimator is small. We are not aware of an expression for the variance of $\widehat{\mathrm{CVaR}}^{\mathrm{IS}}_\beta[X]$ for fixed n. Therefore, we use the asymptotic result (4.4) and use

$$\frac{\mathbb{V}_\varphi[(X(\cdot) - \mathrm{VaR}_\beta[X])_+\, w(\cdot)]}{n(1-\beta)^2} \tag{4.5}$$

as the "variance" of $\widehat{\mathrm{CVaR}}^{\mathrm{IS}}_\beta[X]$ for fixed n.

4.2. Deriving the optimal biasing distribution. Before we construct the proposed biasing density φ, we first compute the optimal biasing density, i.e., the biasing density that gives an estimator $\widehat{\mathrm{CVaR}}^{\mathrm{IS}}_\beta[X]$ with zero variance in (4.5). Although this optimal biasing density is usually impractical (it depends on the quantities to be estimated), it guides us in the construction of a computable biasing density.

Theorem 4.5. *The biasing density resulting in zero variance in (4.5) is given by*

$$\varphi^*(\xi) = \frac{\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\,(X(\xi) - VaR_\beta[X])\,\rho(\xi)}{(1-\beta)\,(CVaR_\beta[X] - VaR_\beta[X])}. \tag{4.6}$$

Proof. We begin by analyzing the variance term in (4.5), which gives

$$
\begin{aligned}
&\mathbb{V}_\varphi\left[\left(X(\cdot)-\mathrm{VaR}_\beta[X]\right)_+ w(\cdot)\right]\\
&=\mathbb{E}_\varphi\left[\left(X(\cdot)-\mathrm{VaR}_\beta[X]\right)_+^2 w^2(\cdot)\right]-\left(\mathbb{E}_\varphi\left[\left(X(\cdot)-\mathrm{VaR}_\beta[X]\right)_+ w(\cdot)\right]\right)^2\\
&=\int_{\widetilde{\Xi}}\left(\frac{\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)_+\rho(\xi)}{\varphi(\xi)}\right)^2\varphi(\xi)d\xi-\left(\int_{\widetilde{\Xi}}\frac{\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)_+\rho(\xi)}{\varphi(\xi)}\varphi(\xi)d\xi\right)^2\\
&=\int_{\widetilde{\Xi}}\frac{\left(\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\right)^2\rho(\xi)}{\varphi(\xi)}\rho(\xi)d\xi-\left(\int_{\widetilde{\Xi}}\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\rho(\xi)d\xi\right)^2\\
&=\int_{\Xi}\frac{\left(\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\right)^2\rho(\xi)}{\varphi(\xi)}\rho(\xi)d\xi-\left(\int_{\Xi}\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\rho(\xi)d\xi\right)^2\\
&=\int_{\Xi}\frac{\left(\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\right)^2\rho(\xi)}{\varphi(\xi)}\rho(\xi)d\xi-(1-\beta)^2\left(\mathrm{CVaR}_\beta[X]-\mathrm{VaR}_\beta[X]\right)^2,
\end{aligned}
$$

where the last identity follows from (4.1) and (2.10) with $\Pr\left[X=\mathrm{VaR}_\beta[X]\right]=0$. The change from $\widetilde{\Xi}$ to Ξ in the second-to-last inequality is justified, since for $\xi\in\Xi\cap(\widetilde{\Xi})^c$ we have $\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)=0$ (see (4.3)), while for $\xi\in\widetilde{\Xi}\cap(\Xi)^c$ we have that $\rho(\xi)=0$.

Now, define

$$
U(\xi)\coloneqq\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\rho(\xi)-(1-\beta)\left(\mathrm{CVaR}_\beta[X]-\mathrm{VaR}_\beta[X]\right)\varphi(\xi).
$$

Using the definition of U and the same arguments as above, we obtain

$$
\begin{aligned}
&\int_{\widetilde{\Xi}}\frac{(U(\xi))^2}{\varphi(\xi)}d\xi\\
&=\int_{\widetilde{\Xi}}\frac{\left(\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\rho(\xi)\right)^2}{\varphi(\xi)}d\xi\\
&-2(1-\beta)\left(\mathrm{CVaR}_\beta[X]-\mathrm{VaR}_\beta[X]\right)\int_{\widetilde{\Xi}}\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)_+\rho(\xi)d\xi\\
&+(1-\beta)^2\left(\mathrm{CVaR}_\beta[X]-\mathrm{VaR}_\beta[X]\right)^2\int_{\widetilde{\Xi}}\varphi(\xi)d\xi\\
&=\int_{\Xi}\frac{\left(\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\left(X(\xi)-\mathrm{VaR}_\beta[X]\right)\right)^2\rho(\xi)}{\varphi(\xi)}\rho(\xi)d\xi-(1-\beta)^2\left(\mathrm{CVaR}_\beta[X]-\mathrm{VaR}_\beta[X]\right)^2,
\end{aligned}
$$

which shows that

$$
\mathbb{V}_\varphi[\left(X(\cdot)-\mathrm{VaR}_\beta[X]\right)_+ w(\cdot)]=\int_{\widetilde{\Xi}}\frac{(U(\xi))^2}{\varphi(\xi)}d\xi.
$$

The density $\varphi=\varphi^*$ in (4.6) gives $U\equiv 0$. ∎

4.3. Using a ROM to generate the biasing density. The optimal density φ^* depends on quantities $\mathrm{CVaR}_\beta[X]$ and $\mathrm{VaR}_\beta[X]$ that we want to estimate; it requires knowledge of the true risk region $\mathbb{G}_\beta[X]$, and it requires evaluating X. Thus it is impractical to use; see also [9]. However, it guides us in the construction of a feasible biasing density, as described next. The optimal biasing density (4.6) motivates the initial choice

$$\varphi(\xi) = \frac{\mathbb{I}_{\mathbb{G}_\beta[X]}(\xi)\,\rho(\xi)}{1-\beta}. \tag{4.7}$$

(Note that $\Pr[\mathbb{G}_\beta[X]] = 1-\beta$ under our assumption that the c.d.f. $H_X(x) = \Pr[X \le x]$ is continuous at $x = \mathrm{VaR}_\beta[X]$.) This choice (4.7) is obtained from the optimal density (4.6) by dropping $\mathrm{CVaR}_\beta[X]$, $\mathrm{VaR}_\beta[X]$, and X. This biasing density has the same support as the optimal one, i.e., $\mathrm{supp}(\varphi^*) = \mathrm{supp}(\varphi)$. However, (4.7) still depends on the risk region of the expensive X. Therefore, we use a ROM and the ϵ-risk region (3.4) to construct our biasing density

$$\varphi(\xi) \coloneqq \frac{\mathbb{I}_{\mathbb{G}^\epsilon_\beta[X_r]}(\xi)\,\rho(\xi)}{\Pr[\mathbb{G}^\epsilon_\beta[X_r]]}. \tag{4.8}$$

Since $\mathbb{G}_\beta[X] \subseteq \mathbb{G}^\epsilon_\beta[X_r]$, the support $\widetilde{\Xi} = \mathbb{G}^\epsilon_\beta[X_r]$ of this density satisfies (4.3). A crucial advantage of (4.8) over (4.6) or (4.7) is that its construction requires only inexpensive evaluations of the ROM.

With the density φ in (4.8), the weight function (4.2) is given by

$$w(\xi) = \frac{\Pr\big[\mathbb{G}^\epsilon_\beta[X_r]\big]}{\mathbb{I}_{\mathbb{G}^\epsilon_\beta[X_r]}(\xi)}. \tag{4.9}$$

Samples ξ from the biasing distribution φ in (4.8) satisfy $\xi \in \mathbb{G}^\epsilon_\beta[X_r]$. Therefore, w in (4.9) is well-defined. Furthermore, for these samples,

$$1-\beta \le w(\xi) = \Pr[\mathbb{G}^\epsilon_\beta[X_r]] \le 1.$$

Note that the smaller the ROM error ϵ_r, the closer $w(\xi) = \Pr[\mathbb{G}^\epsilon_\beta[X_r]]$ to $\Pr[\mathbb{G}_\beta[X]] = 1-\beta$.

The goal of IS is that the IS estimator has much lower variance than a standard Monte Carlo estimator (obtained from (4.5) with $w \equiv 1$). We show next that our proposed density indeed reduces the variance in (4.5), and that the variance reduction is proportional to the size of the ϵ-risk region.

Theorem 4.6. *The variance (4.5) corresponding to the $\widehat{CVaR}^{IS}_\beta[X]$ estimator with density (4.8) is reduced by a factor of at least $\Pr[\mathbb{G}^\epsilon_\beta[X_r]]$ compared to the standard Monte Carlo estimator $\widehat{CVaR}^{MC}_\beta[X]$ (obtained from (4.5) with $w \equiv 1$) with original density ρ, i.e.,*

$$\frac{\mathbb{V}_\varphi\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot) - VaR_\beta[X])\,w(\cdot)\right]}{\mathbb{V}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot) - VaR_\beta[X])\right]} \le \Pr\big[\mathbb{G}^\epsilon_\beta[X_r]\big].$$

Proof. Using the definition of the variance $\mathbb{V}_\varphi$ and the definitions (4.2), (4.9) of the weight,

$$\begin{aligned}
&\mathbb{V}_\varphi\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])\,w(\cdot)\right]\\
&=\mathbb{E}_\varphi\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])^2\,w^2(\cdot)\right]-\left(\mathbb{E}_\varphi\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])\,w(\cdot)\right]\right)^2\\
&=\mathbb{E}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])^2\,w(\cdot)\right]-\left(\mathbb{E}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])\right]\right)^2\\
&=\Pr\left[\mathbb{G}^\epsilon_\beta[X_r]\right]\,\mathbb{E}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])^2\right]-\left(\mathbb{E}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])\right]\right)^2\\
&=\Pr\left[\mathbb{G}^\epsilon_\beta[X_r]\right]\,\mathbb{V}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])^2\right]\\
&\quad-\left(1-\Pr[\mathbb{G}^\epsilon_\beta[X_r]]\right)\left(\mathbb{E}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])\right]\right)^2\\
&\le\Pr\left[\mathbb{G}^\epsilon_\beta[X_r]\right]\,\mathbb{V}_\rho\left[\mathbb{I}_{\mathbb{G}_\beta[X]}(\cdot)\,(X(\cdot)-\mathrm{VaR}_\beta[X])^2\right].
\end{aligned}$$

Thus, dividing the variance term gives the stated result. ∎

4.4. Implementation details of the IS approach. To execute our IS approach, we need to be able to sample from the distribution with density (4.8). Although (4.8) only involves an inexpensive ROM, sampling from the distribution with density (4.8) is still impossible to do exactly. There are several options including estimating the ϵ-risk region and then using acceptance-rejection sampling, or estimating the ϵ-risk region and then approximating (4.8) using a Gaussian mixture model or kernel density estimation. We have applied both acceptance-rejection sampling and a Gaussian mixture model and found that in our numerical example, where the original density ρ is constant, acceptance-rejection sampling performed better. Therefore, we describe acceptance-rejection sampling here. Our approach of approximating (4.8) using a Gaussian mixture model and corresponding numerical results is given in the supplementary materials; see supplementary section SM2.

First, we approximate the ϵ-risk region $\mathbb{G}^\epsilon_\beta[X_r] \coloneqq \{\xi : X_r(\xi)+\epsilon_r(\xi) \ge \mathrm{VaR}_\beta[X_r-\epsilon_r]\}$ by taking m samples of the ROM. Next, we employ the acceptance-rejection sampling strategy to generate samples from density (4.8).

The acceptance-rejection algorithm for continuous random variables generates samples from a desired distribution φ given an easy-to-sample distribution ρ (see, e.g., [9, section 4.7]). The acceptance-rejection algorithm assumes the existence of $C>0$ with $\varphi(\xi)/\rho(\xi) \le C$ for all $\xi \in \Xi$. In our case

$$\frac{\varphi(\xi)}{\rho(\xi)} = \frac{\mathbb{I}_{\mathbb{G}^\epsilon_\beta[X_r]}(\xi)}{\Pr[\mathbb{G}^\epsilon_\beta[X_r]]},$$

and the previous assumption is satisfied with $C = \left(\Pr[\mathbb{G}^\epsilon_\beta[X_r]]\right)^{-1}$. Then the standard acceptance-rejection algorithm generates a candidate sample ξ^c from ρ and a sample u from the uniform distribution $U(0,1)$, and accepts the sample ξ^c if

$$u \le \frac{\varphi(\xi^c)}{C\rho(\xi^c)} = \mathbb{I}_{\mathbb{G}^\epsilon_\beta[X_r]}(\xi^c).$$

Thus, in order to accept the sample we just need to check that it belongs to $\mathbb{G}^\epsilon_\beta[X_r]$, which can be done by evaluating the ROM X_r and its error function ϵ_r at ξ^c.

The ROM-based acceptance-rejection method to generate the samples from a distribution with density (4.8) is given in Algorithm 4.1. Finally, the proposed approach to compute CVaR_β via IS is summarized in Algorithm 4.2.

Algorithm 4.1. Acceptance-rejection sampling using surrogate model.

Input: Surrogate model QoI X_r with error function ϵ_r, risk level $\beta \in (0,1)$, nominal distribution ρ, # of samples m to estimate $\mathbb{G}^\epsilon_\beta[X_r]$, and desired # of samples n.
Output: n samples from (4.8) and an estimate of $\Pr[\mathbb{G}^\epsilon_\beta[X_r]]$.

1: Sample m inputs $\{\xi^{(1)}, \ldots, \xi^{(m)}\}$ from the nominal distribution $\rho(\xi)$ with equal probabilities $p^{(j)} \equiv 1/m$.
2: Compute ROM QoI values $X_r(\xi^{(1)}), \ldots, X_r(\xi^{(m)})$.
3: Compute ROM error function values $\epsilon_r(\xi^{(1)}), \ldots, \epsilon_r(\xi^{(m)})$.
4: Apply steps 2, 3, and 4 of Algorithm 2.1 with X replaced by $X_r - \epsilon_r$ to obtain $\widehat{\mathrm{VaR}_\beta}[X_r - \epsilon_r]$.
5: Estimate ϵ-risk region $\widehat{\mathbb{G}^\epsilon_\beta}[X_r] := \{\xi^{(j)} : X_r(\xi^{(j)}) + \epsilon_r(\xi^{(j)}) \geq \widehat{\mathrm{VaR}}_\beta[X_r - \epsilon_r]\}$.
6: Estimate $\Pr[\mathbb{G}^\epsilon_\beta[X_r]]$ as $\widehat{\Pr}[\mathbb{G}^\epsilon_\beta[X_r]] = |\widehat{\mathbb{G}^\epsilon_\beta}[X_r]|/m$, where $|\cdot|$ denotes cardinality of a set.
7: Set $n_a = 0$ (counter of accepted samples), $n_c = 0$ (counter of candidate samples).
8: **while** $n_a \leq n$ **do**
9: Generate candidate sample ξ^c from $\rho(\xi)$.
10: Compute ROM QoI value $X_r(\xi^c)$.
11: Compute ROM error function value $\epsilon_r(\xi^c)$.
12: **if** $X_r(\xi^c) + \epsilon_r(\xi^c) \geq \widehat{\mathrm{VaR}_\beta}[X_r - \epsilon_r]$ **then**
13: Accept sample ξ^c, set $n_a = n_a + 1$.
14: **end if**
15: Set $n_c = n_c + 1$.
16: **end while**

Algorithm 4.2. Estimating CVaR_β using importance sampling with ROM.

Input: FOM QoI X, ROM QoI X_r with error function ϵ_r, risk level $\beta \in (0,1)$, nominal distribution ρ, # of evaluations m of the ROM X_r, and n of the FOM X.
Output: Importance sampling estimate $\widehat{\mathrm{CVaR}}^{\mathrm{IS}}_\beta[X]$.

1: Apply the acceptance-rejection Algorithm 4.1 to generate n samples from the biasing density (4.8) and obtain the estimate $\widehat{\Pr}[\mathbb{G}^\epsilon_\beta[X_r]]$ of $\Pr[\mathbb{G}^\epsilon_\beta[X_r]]$.
2: Compute FOM outputs $X(\xi^{(1)}), \ldots, X(\xi^{(n)})$.
3: Assign values of weight function $w = \rho/\varphi$: $w(\xi^{(1)}), \ldots, w(\xi^{(n)}) \equiv \widehat{\Pr}[\mathbb{G}^\epsilon_\beta[X_r]]$.
4: Apply Algorithm 2.1 with $X(\xi^{(j)})$, $p^{(j)} = w(\xi^{(j)})/n$, $j = 1, \ldots, n$, to obtain $\widehat{\mathrm{CVaR}}^{\mathrm{IS}}_\beta[X]$.

In Algorithm 4.1, n_c is used to keep track of the number of candidate samples generated from ρ. The probability that a candidate sample is accepted is $1/C = \Pr[\mathbb{G}^\epsilon_\beta[X_r]]$; see [9, Theorem 4.2]. Thus the ratio of the number n of accepted samples to the number n_c of candidate samples is $n/n_c \approx \Pr[\mathbb{G}^\epsilon_\beta[X_r]]$ for large sample sizes. Thus, using $\widehat{\Pr}[\mathbb{G}^\epsilon_\beta[X_r]]$ from

line 6 of Algorithm 4.1, we estimate the number of candidate samples needed as

$$n_c \approx n/\widehat{\Pr}[\mathbb{G}_\beta^\epsilon[X_r]].$$

The total cost of generating n biased samples is approximately

$$m + n/\widehat{\Pr}[\mathbb{G}_\beta^\epsilon[X_r]] \leq m + n/\widehat{\Pr}[\mathbb{G}_\beta[X_r]] = m + \frac{n}{1-\beta} \tag{4.10}$$

ROM evaluations.

5. Numerical results.

We present numerical results for our approach for estimating CVaR_β using ROMs. Section 5.1 introduces the PDE model, followed by a detailed description of its discretization and reduced-order modeling in section 5.2. The numerical results are presented and discussed in sections 5.3 and 5.4.

5.1. Convection-diffusion-reaction PDE model.

We consider a simplified model of a premixed combustion flame at constant and uniform pressure, and follow the notation and set up in [1, sect. 3]. The model includes a one-step reaction of the species

$$2H_2 + O_2 \ \rightarrow 2H_2O$$

in the presence of an additional nonreactive species, nitrogen. The state is comprised of the components $y = [T, Y_F, Y_O, Y_P]$, with the Y_i being the mass fractions of the species fuel (F, here H_2), oxidizer (O, here O_2), product (P, here H_2O), and T denoting the temperature. The physical combustor domain is 1.8 cm in length and 0.9 cm in height. Thus $\Omega = (0, 1.8) \times (0, 0.9)$. Dirichlet boundary conditions are specified on the left boundary $\Gamma_D = \{0\} \times [0, 0.9]$, and homogeneous Neumann boundary conditions are specified in the top, bottom, and right boundary $\Gamma_N = \partial\Omega \setminus \Gamma_D$. The velocity field U is set to be constant in the x_1 direction and divergence free. The molecular diffusivity κ is modeled as constant, equal, and uniform for all species and temperatures. Constants in the nonlinear reaction term are random variables with values in Ξ. For a given $\xi \in \Xi$ the state equation is the system of diffusion-advection-reaction PDEs,

$$0 = \kappa\Delta y(x) - U(x)\nabla y(x) + \mathcal{N}(y(x), \xi), \qquad x \in \Omega, \tag{5.1a}$$

$$y(x)|_{\Gamma_D} = y_D(x), \qquad x \in \Gamma_D, \tag{5.1b}$$

$$\nabla y(x) \cdot \boldsymbol{n} = 0, \qquad x \in \Gamma_N. \tag{5.1c}$$

The left boundary Γ_D is divided into three equal parts. The bottom and top third of the left boundary are held at $T = 300\,\mathrm{K}$ while the mass fractions are prescribed as zero Dirichlet conditions. The middle third of the left boundary is the inflow boundary, where the incoming unburned mixture has temperature $T = 950\,\mathrm{K}$ and mass fractions $Y_{H_2} = 0.0282, Y_{O_2} = 0.2259, Y_{H_2O} = 0$. The nonlinear reaction term $\mathcal{N}(y, \xi) = [\mathcal{N}_T, \mathcal{N}_F, \mathcal{N}_O, \mathcal{N}_P](y, \xi)$ is of Arrhenius type and modeled as

$$\mathcal{N}_i(y, \xi) = -\nu_i \left(\frac{W_i}{\rho}\right) \left(\frac{\rho Y_F}{W_F}\right)^{\nu_F} \left(\frac{\rho Y_O}{W_O}\right)^{\nu_O} A \exp\left(-\frac{E}{RT}\right), \quad i = F, O, P, \tag{5.2a}$$

$$\mathcal{N}_T(y, \xi) = Q \cdot \mathcal{N}_P(y, \xi). \tag{5.2b}$$

Table 5.1
Parameters for the PDE model (5.1)–(5.2) from [1].

Parameter	Physical meaning	Value
κ	molecular diffusivity	$2\,\mathrm{cm}^2/\mathrm{s}$
U	velocity	$50\,\mathrm{cm/s}$
W_{H_2}	molecular weight	$2.016\,\mathrm{g/mol}$
W_{O_2}	molecular weight	$31.9\,\mathrm{g/mol}$
W_{H_2O}	molecular weight	$18\,\mathrm{g/mol}$
ρ	density of mixture	$1.39 \times 10^{-3}\,\mathrm{g/cm}^3$
R	univ. gas constant	$8.314\,\mathrm{J\,mol/K}$
Q	heat of reaction	$9800\,\mathrm{K}$

The uncertain parameters of the model are considered to be

$$\xi = [A, E],$$

with values in the parameter domain

$$\Xi = [A_{\min}, A_{\max}] \times [E_{\min}, E_{\max}] = [5.5 \times 10^{11},\ 1.5 \times 10^{13}] \times [1.5 \times 10^{3},\ 9.5 \times 10^{3}]. \tag{5.3}$$

The random variable $\boldsymbol{\xi}$ is uniformly distributed, i.e., ρ is constant. The other parameters are defined in Table 5.1. For a schematic of the domain and boundary conditions, as well as typical solution behavior, we refer the reader to [1, sect. IV.A].

5.2. Discretization and reduced-order models. This section discuses the discretization of the PDE (5.1), which is our FOM, the QoI, and the computation of ROMs.

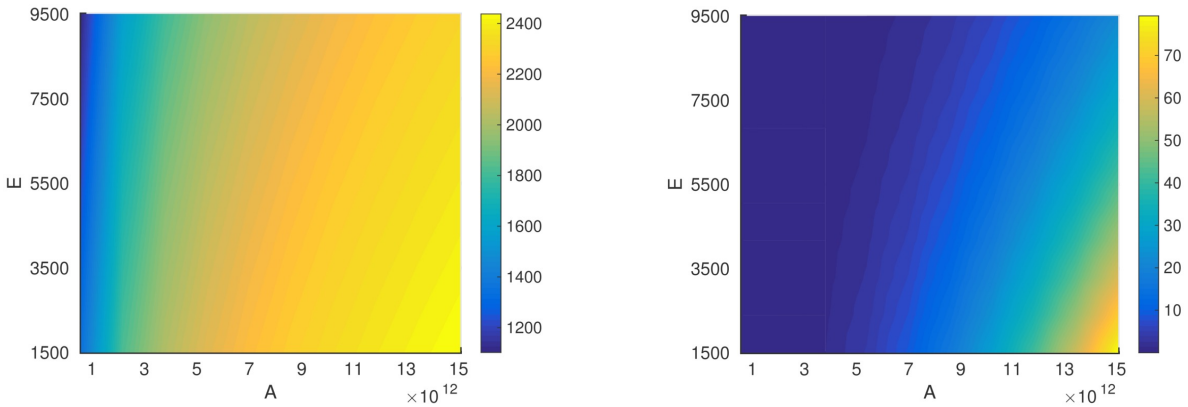
5.2.1. Full-order model. The PDE model is discretized using a finite difference approximation in two spatial dimensions, with a 72×36 grid, leading to $N = 10{,}804$ unknowns in the discretized model. Let $\mathbf{y}$ be the vector with components corresponding to the approximations of the state y at the grid points. The resulting nonlinear system becomes

$$\mathbf{0} = \mathbf{K}\mathbf{y} + \mathcal{N}(\mathbf{y}; \xi), \tag{5.4}$$

with boundary conditions as described above. Here, $\mathbf{K} \in \mathbb{R}^{N \times N}$ is the discretized representation of the derivative operators. The nonlinear system is solved with Newton's method. Let $\mathbf{T}(\cdot) \in \mathbb{R}^{N/4}$ be the vector with components corresponding to the approximations of the temperature $T(x, \cdot)$ at the grid points. Given the uncertainty in the input parameters, the QoI is the random variable

$$X : \Xi \mapsto \mathbb{R}, \qquad X(\xi) = \exp\left(\frac{\|\mathbf{T}(\xi)\|_\infty - 2000}{100}\right). \tag{5.5}$$

This nondimensionalized QoI represents a penalty on temperatures exceeding 2000 K, such as might be imposed when there is a design target but not a hard constraint. For this example, the maximum temperature is between 1122 K and 2435 K, and therefore, the QoI is between 1.5×10^{-4} and 7.8×10^{1}. The maximum temperature $\|\mathbf{T}(\xi)\|_\infty$ and the QoI $X(\xi)$ at parameters in Ξ are shown in Figure 5.1.



(a) Values of $\|\mathbf{T}(\xi)\|_\infty$. (b) Values of $X(\xi)$.

Figure 5.1. *Maximum temperature* $\|\mathbf{T}(\xi)\|_\infty$ *in* K *and the QoI* $X(\xi)$ *for* $\xi = (A, E) \in \Xi$*, computed using the FOM.*

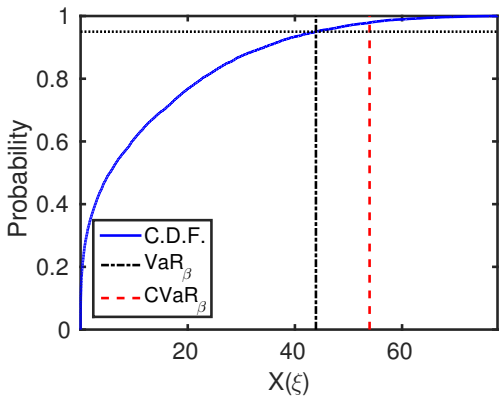


Figure 5.2. *The c.d.f. of the quantity of interest* (5.5) *evaluated with* $n = 10^4$ *samples of the FOM, together with the values of* $VaR_\beta[X]$ *and* $CVaR_\beta[X]$ *for* $\beta = 0.95$.

The c.d.f. of the QoI, together with $\mathrm{VaR}_\beta[X]$ and $\mathrm{CVaR}_\beta[X]$ for $\beta = 0.95$, are shown in Figure 5.2. These quantities are estimated with 10^4 samples of the FOM and are included for illustration only. Here, $\mathrm{VaR}_\beta[X] = 43.92$ and $\mathrm{CVaR}_\beta[X] = 53.94$.

Figure 5.2 and $\mathrm{CVaR}_\beta[X]$ are generated from a large number of FOM samples for illustration only. The goal of this paper is to show that $\mathrm{CVaR}_\beta[X]$ can be estimated with far fewer FOM evaluations using ROMs. Next we describe how these ROMs are computed for this example.

5.2.2. Reduced-order models. We use Proper Orthogonal Decomposition (POD) to compute ROMs of the form

$$\mathbf{0} = \mathbf{K}_r \mathbf{y}_r + \mathcal{N}_r(\mathbf{y}_r; \xi),$$

where the subscript r is a label of the ROM. Here, $\mathbf{K}_r = V_r^T \mathbf{K} V_r$, $\mathbf{y} \approx V_r \mathbf{y}_r$, and we also project the boundary conditions. The projection matrix $V_r \in \mathbb{R}^{N \times N_r}$, $N_r \ll N$, is computed via singular value decomposition of snapshot data $\boldsymbol{Y} = [\mathbf{y}(\xi^{(1)}), \ldots, \mathbf{y}(\xi^{(S)})] \in \mathbb{R}^{N \times S}$, where

$S = 100$ snapshots were generated from solutions of (5.4) at 10×10 equally spaced values A and E in Ξ. Since the nonlinearity is of exponential type, we use the Discrete Empirical Interpolation Method (DEIM) [3] for an efficient evaluation of the nonlinear term. Four different surrogate models are built from $r = 1, 2, 3, 4$ POD basis vectors and the same number of DEIM selection points, respectively. A detailed description of the model reduction for this example is given in [1]. The surrogate models then define a new random variable for the (nondimensional) quantity of interest, namely

$$X_r : \Xi \mapsto \mathbb{R}, \qquad X_r(\xi) = \exp\left(\frac{\|\mathbf{T}_r(\xi)\|_\infty - 2000}{100}\right),$$

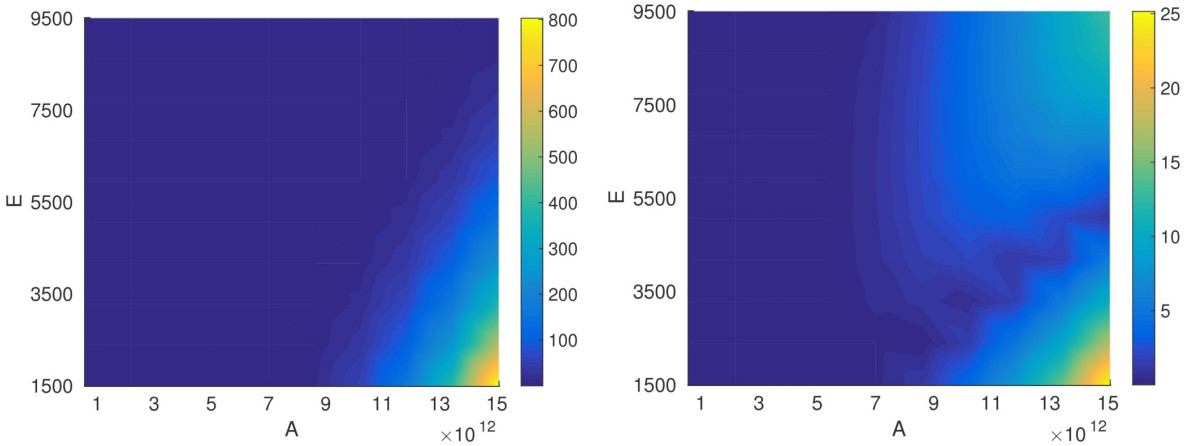
where $\mathbf{T}(\xi)$ is the first block of length $N/4$ in $V_r\mathbf{y}_r$.

Estimates of the errors $\epsilon_r(\xi)$ in the QoI for the four ROMs are shown in Figure 5.3 (note the different error-bar magnitudes). To generate these plots, we compute $\epsilon_r(\xi)$ exactly for parameter values ξ on the 10×10 parameter grid used to generate the ROMs, and then linearly interpolate between these values. Figure 5.3 shows that the maximum errors vary significantly for these four ROMs. In particular, the error in ROM 1 is huge relative to the size of the QoI X; see Figures 5.1 and 5.3a. This is due to the structure of the QoI (5.5). In our case, the ROM maximum temperature exceeds the FOM maximum temperature $\|\mathbf{T}(\xi)\|_\infty$ in the bottom right corner of the parameter domain, where $\|\mathbf{T}(\xi)\|_\infty > 2000\,\mathrm{K}$. This difference is magnified by the exponential function. For example, if the FOM $\|\mathbf{T}(\xi)\|_\infty = 2400\,\mathrm{K}$ and the ROM approximation exceeds this value by only 10%, then $X_r(\xi) = e^{6.4} \approx 601.9$ exceeds $X(\xi) = e^4 \approx 54.6$ by a factor of 11.

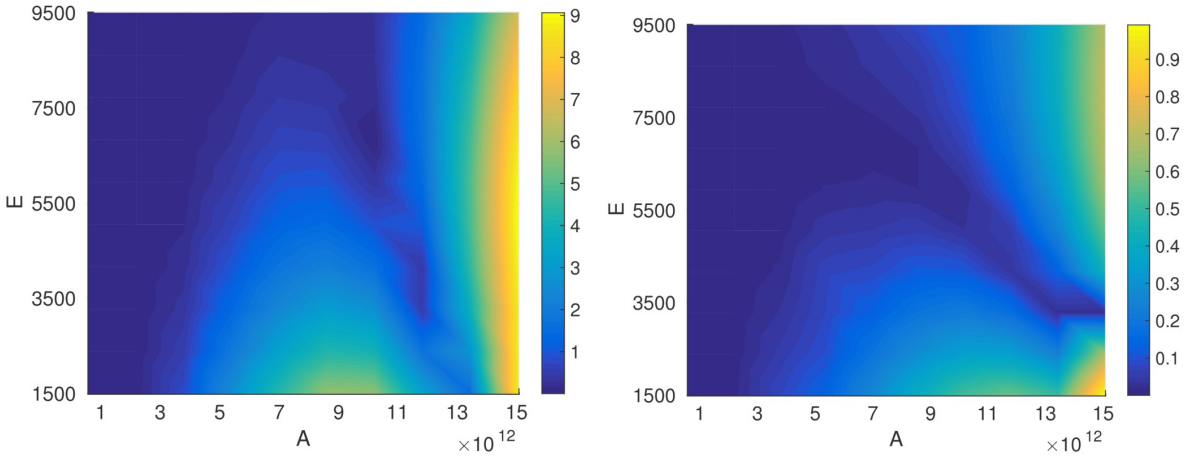
5.3. Estimating CVaR_β by sampling from ROMs. Following our algorithm and error analysis in section 3, we use the ROMs X_r with $r = 1, 2, 3, 4$ basis functions to estimate $\mathrm{CVaR}_\beta[X_r]$ with $\beta = 0.95$. Section 5.3.1 presents the computed risk regions for the FOM and the ROM. Section 5.3.2 computes further ROM errors that occur in the bound for the CVaR_β approximation. Section 5.3.3 then presents the results for CVaR_β estimation.

5.3.1. Risk regions of FOM and ROM. Estimates of the FOM risk region, of the ROM risk regions, and of the ϵ-risk regions are shown in Figure 5.4. These regions are computed using five independent batches of 10^4 samples of the parameters A and E—one batch for the FOM and one batch for each ROM. In every case the ROM risk region is in the bottom right part of the domain Ξ. However, the sizes of the ROM risk regions vary. A sufficiently accurate ROM always gives a good approximation of the true risk region, as is the case with the finest ROM $r = 4$. As expected from Lemma 3.2, the estimated ϵ-risk regions $\widehat{\mathbb{G}}^\epsilon_\beta[X_r]$ contain the estimated FOM risk region $\widehat{\mathbb{G}}_\beta[X]$ and shrink as the ROM becomes more accurate. Again, note that the large number of samples ($n = 10^4$) of the FOM is used only for illustration of the FOM risk region and the error estimates derived in this paper. However, these FOM samples are not used in our ROM or ROM-IS approach to estimate CVaR_β.

5.3.2. ROM errors in risk-regions. Table 5.2 shows the estimate of the error $\widehat{\epsilon}_r^{\max}$ between the FOM and the ROMs in all of Ξ (see (3.2)), the error $\widehat{\epsilon}_r^{G,\mathrm{low}}$ between the FOM and the ROM in the ROM risk region $\widehat{\mathbb{G}}_\beta[X_r]$ (see (3.6)), and the error $\widehat{\epsilon}_r^{G}$ between the FOM and the ROM in the ϵ-risk region $\widehat{\mathbb{G}}^\epsilon_\beta[X_r]$ (see (3.6)). To compute these estimates we use the linear



(a) Error of ROM 1, $\epsilon_1(\xi)$. (b) Error of ROM 2, $\epsilon_2(\xi)$.

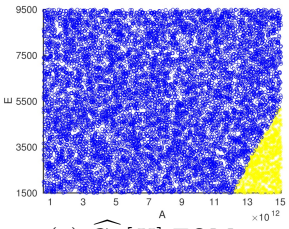


(c) Error of ROM 3, $\epsilon_3(\xi)$. (d) Error of ROM 4, $\epsilon_4(\xi)$.

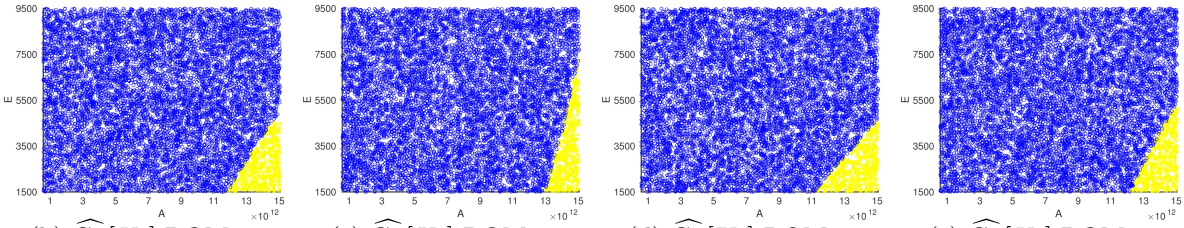
Figure 5.3. *Absolute errors between the QoI X (5.5) and the ROMs X_r for $r = 1, 2, 3, 4$. Note the different magnitudes of the error bars.*

interpolant of the error function $\epsilon_r(\xi)$ displayed in Figure 5.3, evaluate it at the same 10^4 random parameter values that were used to obtain the risk regions, and compute errors as in Algorithm 3.1. In this example, the maximum errors occur in the risk region (compare columns two and four of Table 5.2).

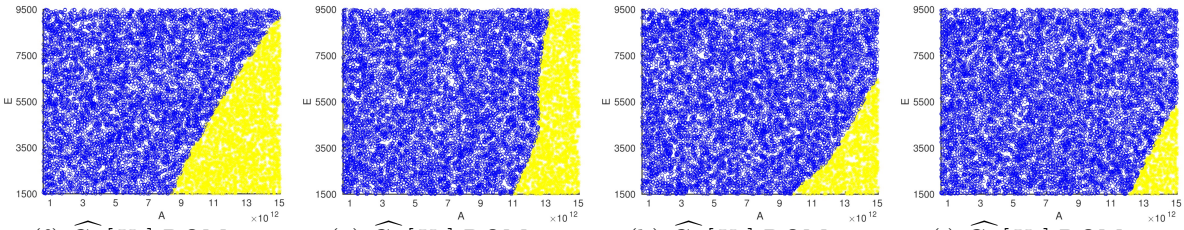
5.3.3. CVaR$_\beta$ estimates with ROM and FOM. Table 5.3 shows the CVaR$_\beta$ estimates with $\beta = 0.95$ for the FOM and the ROMs. These estimates were obtained using Algorithm 2.1 with the $n = 10^4$ samples shown in Figure 5.4a (for FOM) and in Figures 5.4b–5.4e (for ROMs). We denote these CVaR$_\beta$ estimates by $\widehat{\text{CVaR}}_\beta^{\text{MC}}$ to distinguish them from the estimates obtained with IS in the next section. The third column shows the radius of the 95% confidence interval of the respective estimate. We give more details on this computation below. The radius of the 95% confidence interval is a measure of the error between the MC



(a) $\widehat{\mathbb{G}_\beta}[X]$ FOM.



(b) $\widehat{\mathbb{G}_\beta}[X_r]$ ROM 1. (c) $\widehat{\mathbb{G}_\beta}[X_r]$ ROM 2. (d) $\widehat{\mathbb{G}_\beta}[X_r]$ ROM 3. (e) $\widehat{\mathbb{G}_\beta}[X_r]$ ROM 4.



(f) $\widehat{\mathbb{G}^\epsilon_\beta}[X_r]$ ROM 1. (g) $\widehat{\mathbb{G}^\epsilon_\beta}[X_r]$ ROM 2. (h) $\widehat{\mathbb{G}^\epsilon_\beta}[X_r]$ ROM 3. (i) $\widehat{\mathbb{G}^\epsilon_\beta}[X_r]$ ROM 4.

Figure 5.4. *Top row: Estimated FOM risk region* $\widehat{\mathbb{G}_\beta}[X]$. *Middle row: Estimated ROM risk regions* $\widehat{\mathbb{G}_\beta}[X_r]$, $r = 1, 2, 3, 4$. *Bottom row: Estimated* ϵ*-risk regions* $\widehat{\mathbb{G}^\epsilon_\beta}[X_r]$ *of ROMs* $r = 1, 2, 3, 4$. *In all plots* $\beta = 0.95$*; samples in the risk regions are shown as yellow stars, all other samples as blue circles.*

Table 5.2

Estimates of errors between the FOM and the ROMs. The maximum error over the entire parameter domain is $\widehat{\epsilon}_r^{max}$, *the maximum error in the ROM risk region* $\widehat{\mathbb{G}_\beta}[X_r]$ *is given by* $\widehat{\epsilon}_r^{G,low}$, *and the maximum error in the* ϵ*-risk region* $\widehat{\mathbb{G}^\epsilon_\beta}[X_r]$ *is* $\widehat{\epsilon}_r^{G}$.

ROM, r	$\widehat{\epsilon}_r^{\max}$	$\widehat{\epsilon}_r^{G,\text{low}}$	$\widehat{\epsilon}_r^{G}$
1	776.00	776.00	776.00
2	24.47	24.47	24.47
3	9.04	8.72	9.04
4	0.96	0.96	0.96

estimate $\widehat{\text{CVaR}}_\beta^{\text{MC}}$ and the true CVaR_β for the given FOM QoI or ROM QoI. The last two columns of Table 5.3 compare $\widehat{\text{CVaR}}_\beta^{\text{MC}}[X_r]$ with $\widehat{\text{CVaR}}_\beta^{\text{MC}}[X]$.

Tables 5.2 and 5.3 show that

$$\left| \widehat{\text{CVaR}}_\beta^{\text{MC}}[X] - \widehat{\text{CVaR}}_\beta^{\text{MC}}[X_r] \right| \le \widehat{\epsilon}_r^{G}$$

Table 5.3

Estimates of $CVaR_\beta$ with $\beta = 0.95$ for FOM and ROMs $r = 1, 2, 3, 4$ (estimated with 10^4 MC samples and Algorithm 2.1). Absolute and relative errors are computed with respect to the FOM estimate $\widehat{CVaR}_\beta^{MC}[X] = 53.94$.

	$\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}$	CI radius	Abs error	Rel error (%)
FOM	53.94	1.09	—	—
ROM 1	361.40	17.89	307.47	570.05
ROM 2	44.80	0.46	9.14	16.94
ROM 3	49.91	0.92	4.02	7.46
ROM 4	53.87	1.04	0.07	0.13

in this example. In general, this error depends on the bound (3.8) (Figure 5.2 indicates that the c.d.f.'s of X and X_r are continuous) as well as on the MC sampling error.

Note that the $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X]$ estimate, computed with 10^4 expensive FOM samples, can be approximated with high accuracy using ROMs that only require 100 FOM evaluations to generate the ROM. In particular, the estimate $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X_3]$ obtained with the second best ROM has an absolute error 4.02 that is close to the CI width 1.09 of the FOM MC estimator. The estimate $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X_4]$, obtained with the best ROM, happens to have an absolute error that is even below the confidence interval (CI) width 1.09 of the FOM MC estimator. Thus, in this instance, a good ROM can reduce the number of FOM evaluations by a factor of 100. Of course, how much the number of FOM evaluations is reduced depends on the details of the comparison. For example, if $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X]$ is estimated using 10^3 expensive FOM samples, then the CI radius of this FOM estimate is 3.5. In this case the number of FOM evaluations is only reduced by a factor of 10, but ROM 4 gives a better estimate (smaller error with respect to the FOM estimate with 10^4 samples, and small CI radius of 1.04). Overall, a FOM estimate of $\mathrm{CVaR}_\beta[X]$ with fewer samples increases the CI radius. Good ROMs can be used to generate good $\mathrm{CVaR}_\beta[X]$ estimates with high confidence.

The confidence interval is derived from the asymptotic results in [6, sects. 2.1, 2.2] (see also Lemma 4.3). Specifically, the $100(1-\alpha)\%$ CI for $\mathrm{CVaR}_\beta[X]$ is

$$\left[\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X] - z_\alpha \frac{\widehat{\kappa}_\beta}{\sqrt{n}}, \widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X] + z_\alpha \frac{\widehat{\kappa}_\beta}{\sqrt{n}}\right], \tag{5.6}$$

where $z_\alpha = \Phi^{-1}(1-\alpha/2)$ with Φ being the c.d.f. of the standard normal variable, and $\widehat{\kappa}_\beta = \widehat{\psi}_\beta/(1-\beta)$ with

$$\begin{aligned}(\widehat{\psi}_\beta)^2 =& \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\widehat{\mathbb{G}}_\beta[X]}(\xi^{(j)}) \left(X(\xi^{(j)}) - \widehat{\mathrm{VaR}}_\beta^{\mathrm{MC}}[X]\right)^2 w(\xi^{(j)})^2 \\ &- \left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\widehat{\mathbb{G}}_\beta[X]}(\xi^{(j)}) \left(X(\xi^{(j)}) - \widehat{\mathrm{VaR}}_\beta^{\mathrm{MC}}[X]\right) w(\xi^{(j)})\right)^2,\end{aligned} \tag{5.7}$$

where

$$\mathbb{I}_{\widehat{\mathbb{G}}_\beta[X]}(\xi) = \begin{cases} 1 & \text{if } X(\xi) \geq \widehat{\mathrm{VaR}}_\beta^{\mathrm{MC}}[X], \\ 0 & \text{else,} \end{cases}$$

and $w(\xi) \equiv 1$.[2] The confidence intervals for $\mathrm{CVaR}_\beta[X_r]$ of ROMs $r = 1, 2, 3, 4$ are constructed in a similar manner. The third column of Table 5.3 reports the radius (half the width) of the corresponding CI with $z_\alpha = 1.96$, which gives a 95% CI.

Note that $(\widehat{\psi}_\beta)^2$ is an estimate for the variance $\mathbb{V}_\rho\left[(X(\cdot) - \mathrm{VaR}_\beta[X])_+\right]$ which appears in Lemma 4.3 and [6, eq. (8)]. We do not have a proof of consistency of $\widehat{\psi}_\beta$, which is beyond the scope of this paper, but refer the reader to the discussion of a similar result regarding the consistency of an estimator of variance for VaR_β, which can be found in [4].

5.4. Estimating CVaR_β via ROM-informed importance sampling. Next, we use importance sampling to estimate $\mathrm{CVaR}_\beta[X]$. Section 5.4.1 shows the computed estimates. By using IS, we reduce the variance of the estimator compared to a standard MC estimator. We discuss variance reduction for this example in section 5.4.2 and show its alignment with the theoretical results.

5.4.1. Importance sampling estimates of CVaR_β. We generate IS samples from the biasing densities as in section 4.4. Since we consider multiple ROMs, we label φ_r as the biasing density corresponding to ROM X_r. The estimates $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]$ with densities φ_r, $r = 1, 2, 3, 4$, are reported in Table 5.4. The estimates are obtained using Algorithm 4.2 with $m = 10^4$ ROM evaluations to explore the risk regions. We use $n_c = n/\widehat{\Pr}[\mathbb{G}_\beta^\epsilon[X_r]] \le n/0.05 = 2 \cdot 10^3$ ROM evaluations in the acceptance-rejection step to get $n = 100$ samples (see (4.10)). Only $n = 100$ FOM evaluations are used to construct the IS estimates. The reference value is $\mathrm{CVaR}_\beta^{\mathrm{ref}} := \widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X] = 53.94$ from 10^4 samples as reported in Table 5.3. We define the 95% CIs as in (5.6) with $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X]$ substituted by $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]$, w given by the IS weight, and $z_\alpha = 1.96$. The reported results are averaged over $K = 100$ independent trials. The presented CVaR_β estimates are the average values over these trials; similarly the radii of the CIs, and the absolute and relative errors are the average values of respective quantities computed for each independent trial. The mean absolute error (MAE) and mean relative error (MRE) are computed as

$$\mathrm{MAE} = \frac{1}{K}\sum_{k=1}^{K}\left|\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}^{(k)}}[X] - \mathrm{CVaR}_\beta^{\mathrm{ref}}[X]\right|, \quad \mathrm{MRE} = \frac{\mathrm{MAE}}{\left|\mathrm{CVaR}_\beta^{\mathrm{ref}}[X]\right|} \times 100, \tag{5.8}$$

where $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}^{(k)}}[X]$ is the estimate obtained on kth trial, and $K = 100$ is the number of trials. As can be seen from Table 5.4, the absolute errors of the averaged estimates are smaller than the presented mean absolute errors. Additionally, we compute the mean-squared error (MSE) of each estimate as follows:

$$\mathrm{MSE} = \sum_{k=1}^{K}\left(\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}^{(k)}}[X] - \mathrm{CVaR}_\beta^{\mathrm{ref}}[X]\right)^2. \tag{5.9}$$

Table 5.3 showed that coarse ROMs lead to CVaR_β estimates with larger errors when substituted for the FOM (as is expected from our theory). However, Table 5.4 shows that ROMs can be used to build effective biasing densities. For example, the absolute errors for

[2] We include w here because (5.7) will be used in the next section with nonconstant IS weight function.

Table 5.4

Estimates $\widehat{CVaR}_\beta^{IS}[X]$ *with* $\beta = 0.95$ *obtained using Algorithm 4.2 with reduced-order models* X_r, $r = 1, 2, 3, 4$. *Here* $m = 10^4$ *ROM samples are used to identify the* ϵ*-risk region, and the IS estimator uses* $n = 100$ *samples. Absolute and relative errors are computed with respect to the FOM estimate from Table 5.3. All values are averaged over* $K = 100$ *trials.*

	Av $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]$	Av CI radius	MAE (5.8)	MRE (%) (5.8)	MSE (5.9)
IS 1	54.02	5.19	1.99	3.70	6.44
IS 2	54.39	4.26	1.59	2.96	4.22
IS 3	53.74	2.89	1.20	2.23	2.17
IS 4	53.94	1.61	0.66	1.22	0.65

$\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X_r]$ with ROMs 1–3 in Table 5.3 are substantially larger than for the importance sampling estimate $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]$ that used ROMs 1–3 to compute biasing densities; see Table 5.4.

Table 5.5

Estimated variance reduction (4.6) computed with 100 *samples for IS densities* $r = 1, 2, 3, 4$. *The rightmost column reports estimated probability of* $\widehat{\mathbb{G}}_\beta^\epsilon[X_r]$.

	$\widehat{\mathbb{V}}_\varphi[\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]]/\widehat{\mathbb{V}}_\rho[\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X]]$	$\widehat{\Pr}[\mathbb{G}_\beta^\epsilon[X_r]]$
IS 1	0.2258	0.2463
IS 2	0.1519	0.1771
IS 3	0.0691	0.0967
IS 4	0.0214	0.0519

5.4.2. Variance reduction via IS. Table 5.5 reports the estimated reduction in variance of IS estimators of CVaR_β compared to MC estimators, as well as the computed upper bound. The variances of $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X]$ and $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}[X]$ are estimated using 100 samples from ρ and φ, respectively, and are averaged over 100 trials. The second column of Table 5.5 shows the resulting ratios. The relative sizes of the ϵ-risk regions, $\Pr[\mathbb{G}_\beta^\epsilon[X_r]]$, are estimated using the ROM samples as described in Algorithm 4.1 and are reported in the third column of Table 5.5. Theorem 4.6 shows that the density (4.8) leads to a reduction of the variance from (4.5) by a factor of $\Pr[\mathbb{G}_\beta^\epsilon[X_r]]$. The observed variance reduction in Table 5.5 is in agreement with the theoretical estimate from Theorem 4.6. Observe that as the ROMs become more accurate, $\epsilon_r \to 0$, and $\widehat{\Pr}[\mathbb{G}_\beta^\epsilon[X_r]] \to 1 - \beta = 0.05$. Moreover, note that using ROM 4 results in a variance reduction by a factor of $1/0.0214 \approx 47$, which shows the strength of the IS approach.

5.5. Benefits of using ROMs–computational budget comparisons. Since the ROM construction requires FOM evaluations, the question is, would we be better off using MC or IS with only the FOM? This sections shows that, in general, the answer is no, it is better to use some FOM evaluations to construct a ROM and then use inexpensive ROM evaluations.

Computing the $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{MC}}[X_r]$ estimates in Table 5.3 required an initial 100 FOM samples to train the ROMs, but afterwards no additional FOM evaluations were needed. In comparison, computing the estimate $\widehat{\mathrm{CVaR}}_\beta^{\mathrm{IS}}$ also required 100 FOM samples to train the ROMs, and then another $n = 100$ FOM samples to sample from the biasing density. (In practice one would not average over many runs, so we do not count the cost of those repetitions.) We next compare the ROM and ROM-IS estimators to an estimator that uses 200 FOM samples to directly

estimate CVaR_β.

Table 5.6 reports the CVaR_β estimate obtained using only FOM evaluations. The MC estimate with 200 samples (MC 200 in Table 5.6) is at best as good as IS 1 from Table 5.4 with respect to MSE, MAE, etc. Therefore, the proposed strategy that includes ROMs to get a biasing distribution is more accurate and computationally efficient.

Table 5.6

Estimates $\widehat{CVaR}_\beta^{MC}[X]$ with $\beta = 0.95$ obtained using $n = 100$ (MC 100) and $n = 200$ (MC 200) samples. The estimate $\widehat{CVaR}_\beta^{IS}[X]$ (IS FOM) also uses $n = 200$ samples, namely 100 samples to build the biasing distribution with the FOM by fitting a Gaussian mixture model (see supplementary section SM2) and then 100 IS samples. All values are averaged over $K = 100$ trials.

	Av $\widehat{\mathrm{CVaR}}_\beta[X]$	Av CI radius	MAE (5.8)	MRE (%) (5.8)	MSE (5.9)
MC 100	52.78	10.45	4.14	7.68	27.11
MC 200	53.39	7.52	3.66	6.78	20.31
IS FOM	57.29	0.02	28.12	52.13	988.57

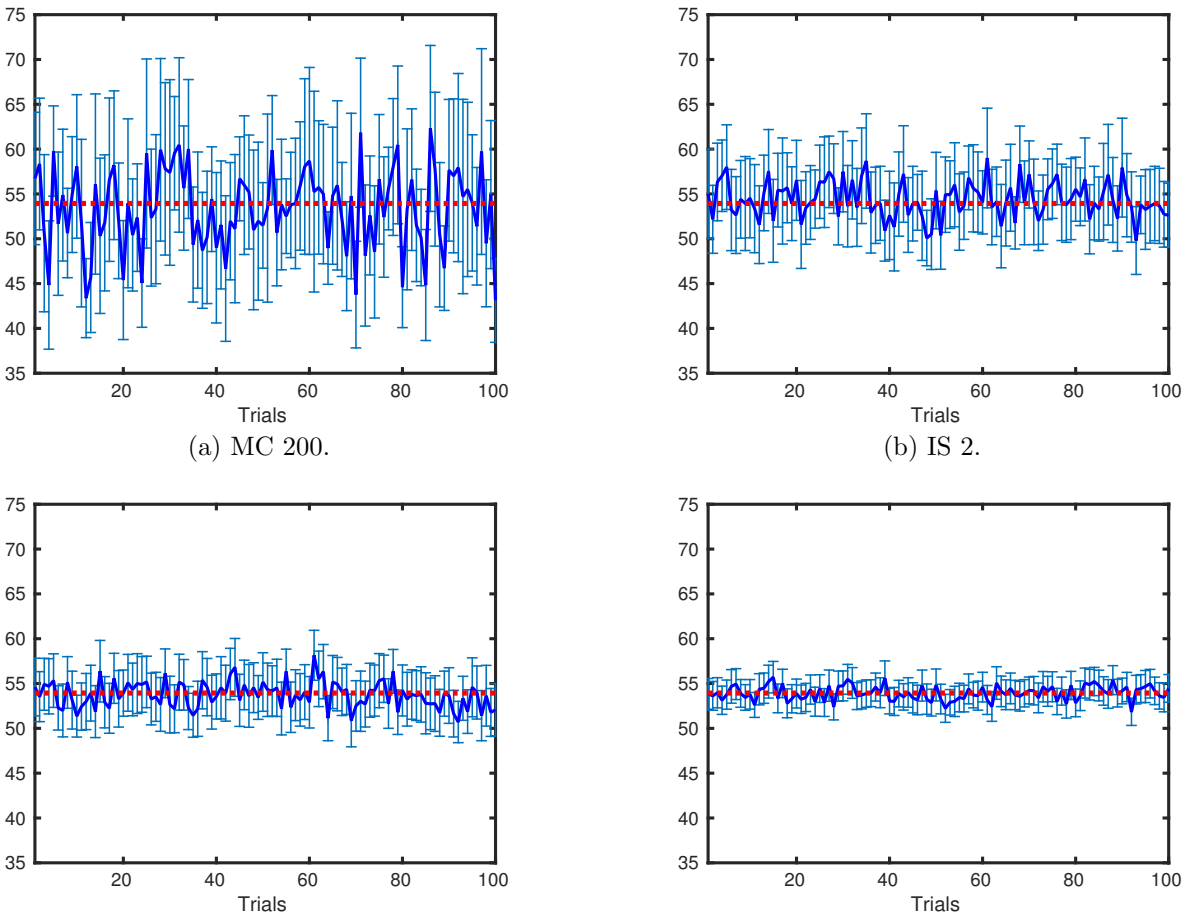
For comparison, Table 5.6 also reports the result of IS with FOM evaluations only. Acceptance-rejection sampling using only FOM evaluations is too expensive: given an estimate of the risk region, we would need $n_c \approx n/(1-\beta) = 2 \cdot 10^3$ FOM evaluations to generate $n = 100$ samples. Therefore, we use IS with the Gaussian mixture model. Specifically, we use 100 MC FOM samples—with only five samples falling into the risk region—and fit a biasing density using a Gaussian mixture model (see supplementary section SM2). We then sample 100 parameters from the obtained density and report the resulting estimate as IS FOM in Table 5.6. This estimate has the largest error of all, mostly due to the small number of samples used to fit the mixture model. Therefore, it appears beneficial to invest the 100 FOM samples to build a ROM and follow the importance sampling strategy described above.

Figure 5.5 presents several of the 100 trials averaged in Tables 5.4 and 5.6 in more detail. It supports the previous observation that investing 100 FOM samples to build a ROM and then generate many inexpensive ROM samples to build a biasing density is beneficial since it substantially reduces the variation of the resulting estimate, given an overall budget of FOM evaluations.

6. Conclusions.

We have presented two methods to estimate CVaR_β with the help of reduced-order models, together with analyses of their performances. One method directly works with the ROM, and the second approach uses the IS framework to reduce the number of high-fidelity samples needed for CVaR_β estimation.

For the first approach in section 3 we showed that the CVaR_β estimation error when using a ROM instead of the high-fidelity model is proportional to the ROM error in the ϵ-risk region of the ROM. Since the ϵ-risk region is small relative to the entire parameter region, this can improve the CVaR_β error estimate. For the second approach, we derived the optimal biasing distribution for the IS framework and used it to derive a biasing distribution that is computed from using only ROM information. We proved that the variance resulting from the proposed biasing distribution is reduced at least by a factor equal to the probability of the risk-region. This factor is small and is asymptotically (as the ROM error goes to zero) equal to $1 - \beta \ll 1$.



(a) MC 200. (b) IS 2.

(c) IS 3. (d) IS 4.

Figure 5.5. *Estimates* $\widehat{CVaR}_\beta^{IS^{(k)}}[X]$ *for* $k = 1, \ldots, 100$ *with* 95% *CIs. The red dotted line corresponds to* $CVaR_\beta^{ref}[X]$.

Both approaches were applied to CVaR_β estimation of a quantity of interest related to heat release modeled by a system of diffusion-advection-reaction PDEs derived from a simple combustion model. For this example, our ROM CVaR_β estimation error from section 3 substantially reduced the computational cost (measured in FOM evaluations). In one comparison the number of FOM evaluations is reduced by a factor of 100. The IS framework led to substantially better CVaR_β estimates when coarser ROMs are used compared to simply replacing FOM samples by ROM samples, but at the expense of using 10^2 additional FOM samples.

For this particular example, we make the following observations: With the same budget of 200 total FOM evaluations, the MC estimates based on the FOM is at best as good as the IS estimate computed with the coarsest ROM 1; more accurate ROMs improve the estimation and especially lead to smaller confidence intervals. With the most accurate ROM 4, our importance sampling framework reduced the variance of the CVaR_β estimator by a factor of

about 47 compared to the standard MC estimator. Overall, our numerical results showed that it appears beneficial to invest FOM samples to train a ROM—which is then used to compute a biasing distribution—than to sample from the FOM directly.

The results in this paper point to future work. Currently, we generate ROMs once from a given number of high-fidelity model evaluations. Since we need only the ROM error to be small in the small ϵ-risk region one could alternate ROM improvement and ϵ-risk region estimation to adaptively generate ROMs and overall use fewer high-fidelity model evaluations. This also might allow increasing the uncertain parameter dimension, as ROM training can be confined to small regions in parameter space.

REFERENCES

[1] M. Buffoni and K. Willcox, *Projection-based model reduction for reacting flows*, in Proceedings of the 40th Fluid Dynamics Conference and Exhibit, 2010, 5008, https://doi.org/10.2514/6.2010-5008.

[2] C. Cannamela, J. Garnier, and B. Iooss, *Controlled stratification for quantile estimation*, Ann. Appl. Stat., 2 (2008), pp. 1554–1580, https://doi.org/10.1214/08-AOAS186.

[3] S. Chaturantabut and D. C. Sorensen, *Nonlinear model reduction via discrete empirical interpolation*, SIAM J. Sci. Comput., 32 (2010), pp. 2737–2764, https://doi.org/10.1137/090766498.

[4] F. Chu and M. K. Nakayama, *Confidence intervals for quantiles when applying variance-reduction techniques*, ACM Trans. Model. Comput. Simul., 22 (2012), 10, https://doi.org/10.1145/2133390.2133394.

[5] M. M. Harajli, R. T. Rockafellar, and J. O. Royset, *Importance sampling in the evaluation and optimization of buffered failure probability*, in Proceedings of the 12th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP12), Vancouver, Canada, 2015, https://doi.org/10.14288/1.0076214.

[6] L. J. Hong, Z. Hu, and G. Liu, *Monte Carlo methods for value-at-risk and conditional value-at-risk: A review*, ACM Trans. Model. Comput. Simul., 24 (2014), 22, https://doi.org/10.1145/2661631.

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

[8] J. Li, J. Li, and D. Xiu, *An efficient surrogate-based method for computing rare failure probability*, J. Comput. Phys., 230 (2011), pp. 8683–8697, https://doi.org/10.1016/j.jcp.2011.08.008.

[9] A. B. Owen, *Monte Carlo Theory, Methods and Examples*, 2013, available online at http://www-stat.stanford.edu/~owen/mc (accessed Nov 22, 2017).

[10] B. Peherstorfer, T. Cui, Y. M. Marzouk, and K. Willcox, *Multifidelity importance sampling*, Comput. Methods Appl. Mech. Engrg., 300 (2016), pp. 490–509, https://doi.org/10.1016/j.cma.2015.12.002.

[11] B. Peherstorfer, B. Kramer, and K. Willcox, *Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models*, Journal of Computational Physics, 341 (2017), pp. 61–75, https://doi.org/10.1016/j.jcp.2017.04.012.

[12] B. Peherstorfer, B. Kramer, and K. Willcox, *Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 737–761, https://doi.org/10.1137/17M1122992.

[13] G. C. Pflug, *Some remarks on the value-at-risk and the conditional value-at-risk*, in Probabilistic Constrained Optimization, Nonconvex Optim. Appl. 49, Kluwer Acad. Publ., Dordrecht, 2000, pp. 272–281, https://doi.org/10.1007/978-1-4757-3150-7_15.

[14] R. T. Rockafellar and J. O. Royset, *Engineering decisions under risk averseness*, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 1 (2015), 04015003, https://doi.org/10.1061/AJRUA6.0000816.

[15] R. T. Rockafellar and S. Uryasev, *Optimization of conditional value-at-risk*, The Journal of Risk, 2 (2000), pp. 21–41, https://doi.org/10.21314/JOR.2000.038.

[16] R. T. Rockafellar and S. Uryasev, *Conditional value-at-risk for general loss distributions*, Journal of Banking & Finance, 26 (2002), pp. 1443–1471, https://doi.org/10.1016/S0378-4266(02)00271-6.

[17] J. O. Royset, L. Bonfiglio, G. Vernengo, and S. Brizzolara, *Risk-adaptive set-based design and applications to shaping a hydrofoil*, Journal of Mechanical Design, 139 (2017), 101403, https://doi.org/10.1115/1.4037623.

[18] A. Ruszczyński and A. Shapiro, *Optimization of convex risk functions*, Math. Oper. Res., 31 (2006), pp. 433–452, https://doi.org/10.1287/moor.1050.0186, see also [19].

[19] A. Ruszczyński and A. Shapiro, *Corrigendum to: "Optimization of convex risk functions" [Math. Oper. Res.,* 31 (2006)*, pp.* 433–452*]*, Math. Oper. Res., 32 (2007), pp. 496, https://doi.org/10.1287/moor.1070.0265.

[20] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming. Modeling and Theory*, MPS/SIAM Ser. Optim. 9, 2nd ed., SIAM, Philadelphia; Mathematical Programming Society (MPS), 2014.

[21] H. Yang and M. Gunzburger, *Algorithms and analyses for stochastic optimization for turbofan noise reduction using parallel reduced-order modeling*, Comput. Methods Appl. Mech. Engrg., 319 (2017), pp. 217–239, https://doi.org/10.1016/j.cma.2017.02.030.

[22] Z. Zou, D. P. Kouri, and W. Aquino, *An adaptive sampling approach for solving PDEs with uncertain inputs and evaluating risk*, in Proceedings of the 19th AIAA Non-deterministic Approaches Conference, AIAA SciTech Forum, (AIAA 2017-1325), 2017, https://doi.org/10.2514/6.2017-1325.