Estimation of Risk Measures with Reduced-Order Models

Boris Kramer\textsuperscript{1}, T. Takhtaganov, M. Heinkenschloss, K. Willcox

Department of Aeronautics and Astronautics
Aerospace Computational Design Laboratory
Massachusetts Institute of Technology

Model Reduction of Parametrized Systems IV
April 11, 2018

\textsuperscript{1} Funded by: DARPA EQUiPS program award number UTA15-001067 and Air Force Center of Excellence on Multi-Fidelity Modeling of Rocket Combustor Dynamics, award FA9550-17-1-0195
Need efficient optimization under uncertainty methods for large-scale systems

- Optimization under uncertainty challenging with large-scale models
  - Cost functions require sampling from expensive solutions
  - Often $\mathcal{O}(100) - \mathcal{O}(10,000)$ solves needed

- Penalizing “tail risk” introduces nonlinear cost functions
  - How to sample efficiently (reduce # of samples) to compute cost function?
Risk-averse optimization

\[
\min_{z \in \mathcal{Z}} \ J(z) := R(s(y(\cdot; z))) + \frac{\alpha}{2} \|z\|_2^2,
\]

where \( y(\xi, z) \) is a solution to the \textit{parameterized high-fidelity model}

\[
F(y(\xi, z), \xi, z) = 0, \quad \forall \xi \in \Xi.
\]

- \( y \): high-fidelity solution that depends on the parameter \( \xi = \xi(\omega) \)
- \( \xi = (\xi_1, \ldots, \xi_M) : \Pi \to \Xi := \prod_{i=1}^{M} \subseteq \mathbb{R}^M \), density \( \rho = \prod_{i=1}^{M} \rho_i \)
- \( R(\cdot) \) is a risk measure
- \( z \): deterministic control/design. One must decide on control before observing outcome \( \xi = \xi(\omega) \).

Output quantity of interest:

\[
X(\xi) := s(y(\xi; z))
\]
Conditional-Value-at-Risk $= \beta$-superquantile

**Value-at-Risk at level** $\beta$ ($=\text{quantile level, often } \beta \geq 0.95$):

$$\text{VaR}_\beta[X] = \inf\{t \in \mathbb{R} : \Pr[X > t] < 1 - \beta\}$$

**Conditional-Value-at-Risk at level** $\beta$: [Rockafellar and Uryasev, 2000]

$$\text{CVaR}_\beta[X] = \text{VaR}_\beta[X] + \frac{1}{1 - \beta} \mathbb{E} \left[(X - \text{VaR}_\beta[X])_+\right]$$

- Choose $\mathcal{R}(X) = \text{CVaR}_\beta[X]$
- Penalizes rare outcomes, length of tail matters in $\text{CVaR}_\beta$
- $\text{CVaR}_\beta$ in engineering design:
  - Royset et al., 2017
  - Yang/Gunzburger, 2017
  - Morio, 2012
  - Zou et al., 2017
How to use ROMs to estimate risk measures efficiently

[Heinkenschloss/K./Takhtaganov/Willcox,’17]

1. Numerical test problem
2. Direct sampling from ROM when error bounds are available
3. Using ROMs/surrogates in importance sampling when error is not known
1) Numerical test problem
Test problem: Convection-diffusion-reaction [Buffoni and Willcox, 2010]

- Model includes a one-step reaction of the species $2H_2 + O_2 \rightarrow 2H_2O$
- Inflow of mixture at left boundary

**Uncertain parameters** of the model relate to reaction terms:

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

**Quantity of interest** $X : \Xi \mapsto \mathbb{R}$ related to discretized temperature

$$X(\xi) = \exp\left(\frac{\|T(\xi)\|_{\infty} - 2000}{100}\right).$$

Values of $\|T(\xi)\|_{\infty}$

Values of $X(\xi)$

c.d.f. of $X(\xi)$, $n = 10^4$
Reduced-order models via POD

- POD ROMs with (D)EIM for Arrhenius reaction terms gives approximate solutions \( y \approx V_r y_r \)
- Projection matrix \( V_r \) from \( S = 100 \) snapshots of HFM at \( 10 \times 10 \) equally-spaced values \( A \) and \( E \) in \( \Xi \)
- Four different ROMs from \( r = 1, 2, 3, 4 \) POD basis vectors
- Surrogate models define a new random variable \( X_r : \Xi \mapsto \mathbb{R} \) with

\[
X_r(\xi) = \exp\left(\frac{\|T_r(\xi)\|_\infty - 2000}{100}\right)
\]

Error of ROM 1, \( \epsilon_1(\xi) \)
Error of ROM 2, \( \epsilon_2(\xi) \)
Error of ROM 4, \( \epsilon_4(\xi) \)
2) CVaR$_\beta$ estimation via direct sampling from ROM
CVaR\(_\beta\) computation

**CVaR\(_\beta\) requires sampling in the tail (=risk) region**
- If the c.d.f. \(H_X(x) = \Pr[X \leq x]\) is continuous at \(x = \text{VaR}_{\beta}[X]\),
  then \(\Pr[X = \text{VaR}_{\beta}[X]] = 0\), and

\[
\text{CVaR}_{\beta}[X] = \frac{1}{1 - \beta} \mathbb{E}[X \cdot \mathbb{1}\{X \geq \text{VaR}_{\beta}[X]\}].
\]

**Definition:** The **risk region** corresponding to \(\text{CVaR}_{\beta}[X]\) is given by

\[
\mathcal{G}_{\beta}[X] := \{\xi \mid X(\xi) \geq \text{VaR}_{\beta}[X]\} \subset \Xi
\]

and the corresponding indicator function of the risk region \(\mathcal{G}_{\beta}[X]\) is

\[
\mathbb{I}_{\mathcal{G}_{\beta}[X]}(\xi) := \mathbb{1}\{X(\xi) \geq \text{VaR}_{\beta}[X]\}.
\]
Reduced-order model (ROM) and risk region

Roadmap: Want bound on $|\text{CVaR}_\beta[X] - \text{CVaR}_\beta[X_r]|$

- Given is a HFM $X(\xi)$ and an approximate quantity of interest $X_r(\xi)$
- Assume the availability of a bound (will be relaxed later):

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

Definition and Lemma: The $\epsilon$-risk region corresponding to $\text{CVaR}_\beta[X]$ is given by

$$\mathcal{G}_\beta^\epsilon[X_r] := \{\xi : X_r(\xi) + \epsilon_r(\xi) \geq \text{VaR}_\beta[X_r - \epsilon_r]\},$$

and define

$$\max_{\xi \in \mathcal{G}_\beta[X] \cup \mathcal{G}_\beta^\epsilon[X_r]} \epsilon_r(\xi) \leq \max_{\xi \in \mathcal{G}_\beta^\epsilon[X_r]} \epsilon_r(\xi) =: \epsilon^G_r.$$

The $\epsilon$-risk region covers the true risk region:

$$\mathcal{G}_\beta[X] \subseteq \mathcal{G}_\beta^\epsilon[X_r] \quad \text{and} \quad \mathcal{G}_\beta[X_r] \subseteq \mathcal{G}_\beta^\epsilon[X_r]$$
Risk regions for four different ROMs

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

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

(f) $\widehat{G}_{\beta}'[X_r]$ ROM 1.
(g) $\widehat{G}_{\beta}'[X_r]$ ROM 2.
(h) $\widehat{G}_{\beta}'[X_r]$ ROM 3.
(i) $\widehat{G}_{\beta}'[X_r]$ ROM 4.
New error bound for ROM-estimated $\text{CVaR}_\beta$

**Theorem:** The error between $\text{CVaR}_\beta$ of the full-order model $X$ and $\text{CVaR}_\beta$ of the reduced-order model $X_r$ is bounded as

$$
|\text{CVaR}_\beta[X] - \text{CVaR}_\beta[X_r]| \\
\leq \left(1 + \frac{\max \{ \text{Pr}\{X = \text{VaR}_\beta[X]\}, \text{Pr}\{X_r = \text{VaR}_\beta[X_r]\} \}}{1 - \beta} \right) \epsilon_r^G \\
\leq \left(1 + \frac{1}{1 - \beta} \right) \epsilon_r^G.
$$

If $X$ and $X_r$ have c.d.f.’s that are continuous at $\text{VaR}_\beta[X]$ and at $\text{VaR}_\beta[X_r]$, respectively, then

$$
|\text{CVaR}_\beta[X] - \text{CVaR}_\beta[X_r]| \leq \epsilon_r^G.
$$

- **Only need error** $\epsilon_r^G$ **in the $\epsilon$-risk region** $G^\epsilon_{\beta}[X_r]$
- **We do not need the error function** $\epsilon_r(\xi)$ in all of $\Xi$
Error bound guides model selection

- Estimates of CVaR$_\beta$ at level $\beta = 0.95$
- Maximum error in ROM $\epsilon$-risk region $\hat{G}_\beta^\epsilon[X_r]$: $\hat{\epsilon}_r^G$

<table>
<thead>
<tr>
<th></th>
<th>CVaR$_\beta^{MC}$</th>
<th>Abs error</th>
<th>Rel error (%)</th>
<th>$\hat{\epsilon}_r^G$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFM</td>
<td>53.94</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>ROM 1</td>
<td>361.40</td>
<td>307.47</td>
<td>570.05</td>
<td>776.00</td>
</tr>
<tr>
<td>ROM 2</td>
<td>44.80</td>
<td>9.14</td>
<td>16.94</td>
<td>24.47</td>
</tr>
<tr>
<td>ROM 3</td>
<td>49.91</td>
<td>4.02</td>
<td>7.46</td>
<td>9.04</td>
</tr>
<tr>
<td>ROM 4</td>
<td>53.87</td>
<td>0.07</td>
<td>0.13</td>
<td>0.96</td>
</tr>
</tbody>
</table>

- Note that $|\hat{\text{CVaR}}^{\beta^{MC}}_\beta[X] - \hat{\text{CVaR}}^{\beta^{MC}}_\beta[X_r]| \leq \hat{\epsilon}_r^G$
- Choice of ROM to sample from can be guided through error bound
What if we don’t have a rigorous error bound

\[ |X(\xi) - X_r(\xi)| \leq \epsilon_r(\xi) \]?

4) CVaR \_\beta estimation using ROMs + importance sampling
Importance sampling: A change of measure

- Define $\text{supp}(\rho) := \{\xi \in \Xi \mid \rho(\xi) > 0\}$.
- Let $\varphi$ be another density with $\text{supp}(\rho) \subseteq \text{supp}(\varphi)$.
- For any integrable function $g : \Xi \to \mathbb{R}$ and $w(\xi) := \frac{\rho(\xi)}{\varphi(\xi)}$ we have

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

- For CVaR$_\beta$, perform change of measure, and account for the change by re-weighting:

$$
\text{CVaR}_\beta[X] = \frac{1}{1 - \beta} \int_{\tilde{\Xi}} \mathbb{I}_{\tilde{G}_\beta[X]}(\xi) X(\xi)w(\xi)\varphi(\xi)\, d\xi
$$

- **Assumption:** The support $\tilde{\Xi}$ of the biasing density $\varphi$ satisfies

$$
\tilde{G}_\beta[X] \subset \tilde{\Xi}.
$$
**Optimal biasing density gives zero variance**

**Lemma:** Under certain conditions, \( \widehat{\text{CVaR}}_{\beta}^{\text{IS}}[X] \to \text{CVaR}_\beta[X] \) w.p. 1 as \( n \to \infty \) and

\[
\sqrt{n} \left( \text{CVaR}_{\beta}^{\text{IS}}[X] - \text{CVaR}_\beta[X] \right) \Rightarrow \frac{\left( \nabla_\varphi \left( (X(\cdot) - \text{VaR}_\beta[X])_+ w(\cdot) \right) \right)^{1/2}}{1 - \beta} N(0, 1).
\]

**Theorem:** The optimal biasing density

\[
\varphi^*(\xi) = \frac{\mathbb{I}_{\text{G}_\beta}[X](\xi) \left( X(\xi) - \text{VaR}_\beta[X] \right) \rho(\xi)}{(1 - \beta) \left( \text{CVaR}_\beta[X] - \text{VaR}_\beta[X] \right)}
\]

results in zero “variance”, i.e.,

\[
\frac{\nabla_\varphi \left( (X(\cdot) - \text{VaR}_\beta[X])_+ w(\cdot) \right)}{n(1 - \beta)^2} = 0.
\]

- **Usual problem:** Optimal biasing density depends on \( \text{CVaR}_\beta, \text{VaR}_\beta \) which we want to compute. \( \Rightarrow \) But helps in finding a good biasing density which results in low variance
Approximation of the optimal IS density via ROMs possible

- Optimal biasing density $\varphi^*$ motivates the initial choice
  \[ \varphi_1(\xi) = \frac{1_{G_\beta[X]}(\xi) \rho(\xi)}{1 - \beta}. \]

- Still depends in the risk region of the expensive $X$

- Use a ROM and $\epsilon$-risk region to get
  \[ \varphi(\xi) := \frac{1_{G_\beta^\epsilon[X_r]}(\xi) \rho(\xi)}{\Pr[G_\beta^\epsilon[X_r]]}. \]

- Since $G_\beta[X] \subseteq G_\beta^\epsilon[X_r]$: $\text{supp}(\rho) = \tilde{\Xi} = G_\beta^\epsilon[X_r] \subseteq \text{supp}(\varphi) = \Xi$

**Theorem:** IS with $\varphi$ reduces variance compared to MC sampling with $\rho$ by

\[
\frac{\nabla_\varphi \left[ 1_{G_\beta[X]}(\cdot) (X(\cdot) - \text{VaR}_\beta[X]) w(\cdot) \right]}{\nabla_\rho \left[ 1_{G_\beta[X]}(\cdot) (X(\cdot) - \text{VaR}_\beta[X]) \right]} \leq \Pr[G_\beta^\epsilon[X_r]].
\]
Importance sampling gives accurate CVaR\_β estimates

- $m = 10^4$ ROM evaluations to explore risk region $G^e_\beta[X_r]$
- Acceptance-rejection algorithm to get samples from biasing distribution
- All $\hat{\text{CVaR}}^\text{IS}_\beta[X]$ estimates use $n = 100$ HFM samples, averaged over $K = 100$ runs
- Reference $\text{CVaR}^\text{ref}_\beta = \text{CVaR}^\text{MC}_\beta[X] = 53.94$ from $10^4$ HFM samples

$\text{MAE} = \frac{1}{K} \sum_{k=1}^{K} \left| \hat{\text{CVaR}}^\text{IS}_\beta^{(k)}[X] - \text{CVaR}^\text{ref}_\beta[X] \right|$, \hspace{0.5cm} $\text{MRE} = \frac{\text{MAE}}{|\text{CVaR}^\text{ref}_\beta[X]|} \times 100\%$

- When used with IS, even inaccurate ROMs can give good estimates

<table>
<thead>
<tr>
<th>Av $\hat{\text{CVaR}}^\text{IS}_\beta[X]$</th>
<th>MAE</th>
<th>MRE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS 1</td>
<td>54.02</td>
<td>1.99</td>
</tr>
<tr>
<td>IS 2</td>
<td>54.39</td>
<td>1.59</td>
</tr>
<tr>
<td>IS 3</td>
<td>53.74</td>
<td>1.20</td>
</tr>
<tr>
<td>IS 4</td>
<td>53.94</td>
<td>0.66</td>
</tr>
</tbody>
</table>
As expected, IS reduces the variance

Estimated variance reduction computed with 100 samples for IS densities $r = 1, 2, 3, 4$.

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\nabla}<em>\varphi [\text{CVaR}^{IS}</em>\beta [X]] / \hat{\nabla}<em>\rho [\text{CVaR}^{MC}</em>\beta [X]]$</th>
<th>$\hat{\Pr}[G^\epsilon_\beta [X_r]]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IS 1</td>
<td>0.2258</td>
<td>0.2463</td>
</tr>
<tr>
<td>IS 2</td>
<td>0.1519</td>
<td>0.1771</td>
</tr>
<tr>
<td>IS 3</td>
<td>0.0691</td>
<td>0.0967</td>
</tr>
<tr>
<td>IS 4</td>
<td>0.0214</td>
<td>0.0519</td>
</tr>
</tbody>
</table>

- **Recall theorem:**

$$\frac{\nabla_\varphi \left[ \mathbb{I}_{G_\beta[X]}(\cdot) (X(\cdot) - \text{VaR}_\beta[X]) \right] w(\cdot)}{\nabla_\rho \left[ \mathbb{I}_{G_\beta[X]}(\cdot) (X(\cdot) - \text{VaR}_\beta[X]) \right]} \leq \hat{\Pr}[G^\epsilon_\beta [X_r]].$$

- As ROMs become more accurate, $\epsilon_r \to 0$, and $\hat{\Pr}[G^\epsilon_\beta [X_r]] \to 1 - \beta = 0.05$. 
Review and conclusion

**Review:**
- Showed that using ROMs to sample from for $\text{CVaR}_\beta$ computations is practical + established error bound
- Using ROMs together with importance sampling can produce efficient estimates (ROM only needs to identify risk-region correctly)

**Conclusions:**
- ROM only has to be accurate in $\epsilon$-risk region $\Rightarrow$ currently working on adaptive ROM construction to make method more efficient
- Computationally, investing in a ROM pays off (shown in paper)